"1. Study Title : ""Journey into SPACE: Evidence-based Design of an App to Reduce Digital Addiction""
Rater 1 Rating = 4 (Very Good)
CR Justification: The ChatGPT version comprehensively covers the socio-technical and psychological insights as the REF submission, providing a holistic view of the app's impact. It additionally includes potential extensions of the app's methodology to other digital health interventions.
Rater 2 Rating = 3: Good
CR Justification: The AI version covers most of the major impacts mentioned in the REF version, such as technological and societal benefits. However, the REF submission provided more detailed accounts of specific interventions, including the social norms approach and qualitative user feedback on the SPACE app's features, which were only briefly mentioned in the AI version
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated report broadly covers most significant aspects of impact, such as technological, social, and health dimensions, highlighting behavior-driven app design to reduce DA, user productivity, and user wellbeing. Nevertheless, the ChatGPT version lacks in-depth descriptions of DARG's policy initiatives and international collaborations with technology platforms as shown in the REF version. Additionally, while it touches on educational and health impacts, these are summarized rather than detailed comprehensively.
Rater 4 Rating = 5
CR Justification: The AI-generated impacts covered an extensive range of impact dimensions, including social, health, and psychological benefits, as well as economic implications via productivity improvements. Each domain aligned well with REF's stated outcomes and the intention behind SPACE's design to foster healthier digital habits. Given this comprehensive scope, the AI content thoroughly captures the multi-dimensional impact present in the original case study.

"
"2. Study Title : Using Macroprudential Policies to Reduce the Risk of Financial Crises
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI-generated content covers most of the key dimensions of impact described in the REF submission, including policy influence, economic implications, and the use of tools for managing financial risk. However, the AI version does not delve deeply into the technical details of the tools or the direct stakeholder engagement described in the REF. Overall, the AI version offers a broad yet insightful overview of the key research impacts.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated impacts covered most of the significant dimensions of the original research's impact, including social, economic, and policy implications. It also touched upon long-term stability and innovation aspects in financial tools. However, the AI-generated version lacked the same depth in discussing the development of tools for financial regulators in specific institutional contexts.
Rater 3 Rating = 3: Good
CR Justification: Matched in Both Versions: Both versions highlight the economic, policy, and academic domains of impact, showing a strong alignment in covering multiple significant aspects of the research's influence. REF Only: The REF submission's specificity in describing contributions to individual institutions and direct engagement with financial bodies reflects a comprehensive impact assessment. ChatGPT Only: The ChatGPT version broadly addresses long-term impact and sustainability, though less specific in institutional context, showing it has room to improve on details found in REF.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated version covers the main areas of impact but does so with less depth in specific use cases, such as distinct national-level applications (e.g., in the G7 countries) and granular effects on fiscal policy adjustments. Although the AI text adequately covers social and economic dimensions, technological impacts are less emphasized.

"
"3. Study Title : Utilising the human-canine relationship to support vulnerable people in the criminal justice system
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content captured most of the significant aspects of impact but lacked detail on some of the specific dimensions present in the REF submission. While the AI version addressed the core impacts, including policy, education, and international influence, it missed some of the granular details, such as specific stakeholder engagement strategies and nuanced outcomes like training specifics for law enforcement or the number of officers trained.
Rater 2 Rating = 3
CR Justification: The AI-generated version covers significant aspects of the impact, such as the psychological and emotional benefits provided by therapy dogs to vulnerable witnesses and

victims. It also touches on both national and international adoption. However, the AI version does not delve into all the specific sectors mentioned in the REF submission, such as training and educational impacts, which were well-explored in the REF case study.
Rater 3 Rating = 5 (Excellent)
CR Justification: ChatGPT's version captures the full range of the project's impact dimensions, including social, legal, educational, and health. Both versions reflect a multi-dimensional impact approach, but the AI-generated content goes further by suggesting potential impacts in healthcare settings and broader mental health applications, extending the original study's findings comprehensively across sectors.
Rater 4 Rating = 3: Good
CR Justification: The AI-generated report includes a broad overview of the major dimensions of impact (social, policy, and psychological well-being) but does not capture the full extent of impacts such as the educational and training contributions emphasized in the REF submission. The AI report acknowledges some multidimensional impacts but lacks in-depth focus on capacity-building elements, especially regarding the training of police and legal personnel.

"

"4. Study Title : Embedding responsible practices into business by taking inspiration from the Quakers
Rater 1 Rating = 3 (Good)
CR Justification: The AI version covered most of the critical aspects of impact, including improved governance, decision-making, and stakeholder collaboration. However, it missed several important details that were present in the REF submission. Specifically, it didn't fully capture the breadth of stakeholder engagement and community-focused outcomes, such as the program to support social mobility and disadvantaged groups through initiatives like ""Bake the Difference.""
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated content covered many essential dimensions of the impact, including economic, social, and organizational benefits of QBM. However, some specific examples of sectoral impact mentioned in the REF submission (e.g., detailed improvements in the UN SDG integration by The Quiet Company) were not explored in full depth in the AI version.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI report covers the core impact dimensions extensively, including social, economic, and ethical domains, mirroring the REF submission's focus on embedding QBM within various sectors. The AI content provides a comprehensive overview, mentioning scalability and transferability to different sectors, cross-disciplinary relevance, and ethical considerations. However, some specific stakeholder examples are less granular.
Rater 4 Rating = 4 (Very Good)
CR Justification: The AI-generated content covers most impact areas specified in the REF submission, addressing social, economic, and governance impacts comprehensively. However, it lacks certain nuances such as QBM's impact on inclusivity through silent reflection practices, which are specifically mentioned in the REF submission.

"

"5. Study Title : Hope 4 The Community CIC: Improving Lives of People Living with Long-Term Health Conditions by providing Self-Management Tools
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated version covers a wide array of impacts, particularly focusing on health outcomes, quality of life improvements, and peer-led interventions. It touches on digital adaptations and social enterprise development but lacks the detailed economic impacts (e.g., income generation of £715K) and specific case studies of policy integration, which are thoroughly addressed in the REF submission. The REF version also outlines the impact of creative arts projects on well-being, which is absent in the AI content .
Rater 2 Rating = 3: Good
CR Justification: The REF submission provides detailed evidence on specific trials, including numbers and outcomes from various cohorts (e.g., 56 cancer survivors, 251 colorectal patients), which are essential for evaluating the depth of the research's impact. The AI-generated analysis focuses more on potential pathways but does not always include the same level of clinical trial data and evidence.
Rater 3 Rating = 4
CR Justification: The AI analysis covers most critical aspects, including mental health, policy implications, and economic impact. It also captures depth across multiple dimensions like social and health impacts. However, AI content occasionally misses specific REF details, particularly in the economic aspects and policy depth, where specific statistics and examples provided by REF were not fully matched.
Rater 4 Rating = 3 - Good

CR Justification: The ChatGPT version covers most of the core dimensions highlighted in the REF submission, including health, social, and policy impacts. However, the REF 2021 report provides more comprehensive details on specific partnerships, training programs, and financial outcomes, which are condensed in the AI summary.

"
"6. Study Title : Localisation of Electricity Generation and Use
Rater 1 Rating = 3
CR Justification: The AI-generated content covers most major aspects of the REF case study, including economic, social, and policy impacts. However, it does not fully address the technical achievements of the research, such as the development of Hestia units and their commercialization through Exergy Devices Ltd, nor the environmental impact of reducing carbon emissions through peer-to-peer trading.
Rater 2 Rating = 3 - Good
CR Justification: The AI version covers the majority of impact dimensions such as environmental, technological, and policy domains. However, it lacks detailed insights into specific community and consumer benefits like savings in Bethesda, which were highlighted in the REF submission.
Rater 3 Rating = 3 - Good
CR Justification: While comprehensive in many aspects, the ChatGPT version misses some nuanced elements of the REF submission, particularly around the regulatory and policy challenges encountered. However, it includes a broad coverage of social, economic, and technological impacts.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover most dimensions of the original research comprehensively, addressing economic, technological, and environmental impacts. However, the depth in stakeholder engagement, particularly around localized community engagement and policy influence, lacks the same level of specificity as the REF-submitted version.

"
"7. Study Title : Critical Connections pedagogical model based on multilingualism and digital storytelling boosts language learning and digital skills
Rater 1 Rating = 4: Very Good
CR Justification: ChatGPT-generated impacts cover the major themes from the REF submission, including teacher training, student engagement, and the creation of a digital community. However, the REF submission includes more detailed examples and specifics regarding geographic impact (e.g., locations like Taiwan) and international engagement (e.g., ECML, NRCSE). ChatGPT's version presents a comprehensive narrative but does not always drill down into these particularities.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated report addresses several core areas of the project, including teacher training, digital storytelling events, and the role of multilingualism in education. However, it does not capture all the specifics, such as the detailed geographical spread of the workshops and events or the full scope of community and student engagement.
Rater 3 Rating = 3 - Good
CR Justification: The AI content covers most significant impact dimensions including social, educational, and technological aspects, with attention to the transformational role of multilingual storytelling. However, the REF version includes additional depth on community-based school integration, specific data on workshop attendance, and references to ongoing professional development, offering a more comprehensive view.
Rater 4 Rating = 4 (Very Good)
CR Justification: The AI-generated report covers most dimensions cited in the REF submission, addressing social, educational, and cultural impacts. Additionally, it captures long-term impacts on educational policy and international scalability. However, it does not detail specific events or partnerships as thoroughly as the REF document, slightly limiting its comprehensiveness.

"
"8. Study Title : Helping to Sustain the UK's Independent Film Industry Through an Improved Risk Management Strategy
Rater 1 Rating = 3 (Good)
CR Justification: The AI version provided a broad overview of the research impact across economic, social, technological, and policy domains, covering most major aspects included in the REF submission. However, it did not address certain nuances, such as specific stakeholder involvement or detailed policy shifts that were mentioned in the REF version.
Rater 2 Rating = 3: Good
CR Justification: The AI-generated version covers many key dimensions of research impact,

including economic, social, and policy dimensions. However, the REF version provides a more comprehensive analysis, especially with regard to the role of international stakeholders and nuanced industry collaborations.
Rater 3 Rating = 5 (Excellent)
CR Justification: The ChatGPT version addresses all major impact dimensions listed in the REF submission, including social, economic, and cultural impacts. It incorporates a wide range of areas, showing a comprehensive understanding of the project's outcomes on both a national and international scale. This version demonstrates coverage of both immediate and potential future impacts.
Rater 4 Rating = 3: Good
CR Justification: The AI-generated content captures most major impact dimensions, covering economic, social, and technological aspects. However, while it reflects on public policy influence and data sharing in industry strategy, it lacks in-depth coverage of specific actions taken by the government stakeholders and exact policy influence milestones seen in the REF submission. The original study's multi-layered approach to sector funding mechanisms and specific policy outcomes (e.g., DCMS engagement and BSF data initiatives) is only partly mirrored in the AI output.

"
"9. Study Title : Changing practice and improving wellbeing through immersive vocal art
Rater 1 Rating = 3: Good
CR Justification: The AI version covers most of the significant impacts mentioned in the REF submission, including contributions to academic knowledge, social impacts, and performance studies. However, certain aspects, such as the long-term, direct community involvement, and the detailed outcomes for disabled audiences in artistic participation, are less comprehensively covered.
Rater 2 Rating = 3: Good
CR Justification: The AI version covers major dimensions of the research impact, such as cultural, social, and academic. However, the REF submission is more comprehensive in detailing the specific community and policy impacts, particularly with local schools and disabled children, providing richer context on how those impacts manifested over time.
Rater 3 Rating = 3 (Good)
CR Justification: The AI version provides a broad overview of dimensions, covering social, cultural, and educational impacts, but does not fully capture all aspects of the detailed case study. Areas such as institutional influence, teacher feedback, and qualitative evaluations from users, which were outlined in the REF, are underrepresented in the AI content, affecting comprehensiveness.
Rater 4 Rating = 3
CR Justification: While the AI-generated impacts covered major domains such as social, cultural, and health/well-being, it lacked the depth found in the REF version regarding specific audience engagement levels, such as educational package dissemination and uptake, and detailed quantitative data (e.g., app downloads, venue visits). The REF report provided extensive context on audience interactions and feedback from pilot programs, showcasing comprehensive project reach and depth, which was only moderately covered by the AI-generated version.

"
"10. Study Title : Safewards: Increasing Safety on Psychiatric Inpatient Wards
Rater 1 Rating = 3 (Good)
CR Justification: While the ChatGPT-generated version covers a wide range of impact areas such as social, economic, and health dimensions, it lacks some specific examples and percentages that the REF submission uses to substantiate its claims. Additionally, certain detailed policy implications and stakeholder endorsements are not explored as thoroughly as in the original version. Overall, the ChatGPT version provides a broader but slightly less detailed account.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated content addressed multiple dimensions of impact, such as health, social, and academic impacts. It correctly identified how the Safewards model impacted staff-patient interactions, public policy, and patient outcomes. However, some specific dimensions of implementation, such as details of the website and online resources, international translation efforts, and the creation of a supportive online community for healthcare professionals, were either underemphasized or omitted in the AI version.
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated content covers significant aspects of impact, including academic, social, health, and policy dimensions. However, it does not fully address all intricate layers of the original submission, such as specific endorsement bodies (e.g., NHSE, Royal College of Nursing) and nuanced descriptions of stakeholder engagement (e.g., social media channels, translation efforts).

Rater 4 Rating = 4 - Very Good
CR Justification: The AI-generated content provides a comprehensive view, covering various
dimensions of impact, including social, health, ethical, and economic impacts. It aligns well
with the REF content by addressing immediate patient outcomes and detailing the Safewards
model's adaptability to international contexts. However, it lacks the specific quantitative data
present in the REF submission that highlights particular reductions in containment practices
within certain hospitals.

"
"11. Study Title : Contemporary documentary practices: historical perspective and
interdisciplinary approaches – the International Research Centre for Interactive Storytelling
(IRIS)
Rater 1 Rating = 3
CR Justification: The AI-generated content covers most major aspects of the research impact
described in the REF submission, including academic and social dimensions. However, it misses
some of the finer, detailed examples of interdisciplinary impacts and the broad engagement
across multiple domains described in the REF. For instance, while the AI-generated version
touches on academic and social impacts, it lacks the comprehensive coverage of cultural,
educational, and artistic impacts evident in the original case.
Rater 2 Rating = 3 – Good
CR Justification: The AI-generated content provides a good overview of the project's societal
impacts and engages with multiple research domains (social, cultural, academic), but lacks some
specific details about community interactions and policy influence that are present in the REF
submission. The coverage of both social and technological impacts is adequate but not as
detailed.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI-generated report encompasses a wide spectrum of impacts, covering
social, cultural, technological, and policy impacts similar to those in the REF submission.
While comprehensive, some depth was lacking in specific examples of institutional-level changes,
such as IRIS's influence on international collaboration. Nevertheless, it broadly captures the
key impact domains and pathways described in the REF submission.
Rater 4 Rating = 4: Very Good
CR Justification: The AI-generated content is comprehensive in addressing the academic, social,
cultural, and technological impacts of the study, with a balanced focus on interdisciplinary
influence. However, certain elements, like educational engagement through VR workshops, which
appear in the REF submission, receive limited emphasis.

"
"12. Study Title : Improving patient outcomes through better project management of clinical
trials
Rater 1 Rating = 3
CR Justification: The AI-generated impacts cover most of the significant dimensions (economic,
social, and health), such as cost savings for pharmaceutical companies and improved patient
outcomes. However, some detailed points related to specific collaborations (e.g., Blau
Farmaceutica's incorporation of project management methods into their SOPs) were not mentioned
in the AI-generated version. Additionally, while the AI covers broader concepts, it sometimes
lacks the depth and specificity of examples provided in the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: The AI version captured many key dimensions, such as economic, policy, and
operational impacts. However, it was less comprehensive regarding the in-depth stakeholder
analysis provided in the REF submission. For example, REF 2021 detailed multiple case studies
(LEO Pharma, Blau Farmaceutica) and specific examples of cost savings through the elimination of
change orders. While these were touched upon in the AI-generated content, some details were
omitted, impacting the overall comprehensiveness.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated content addresses all critical dimensions of impact,
including the economic, health, academic, and policy implications of the research, closely
matching the REF submission's multi-dimensional impact approach. It also touches on scalability
and cross-disciplinary potential, although with less specificity about individual pharmaceutical
entities. Despite minor omissions in specific examples, the AI content provides a thorough view
of the research's potential reach and long-term significance, aligning well with the REF details.
Rater 4 Rating = 3 (Good)
CR Justification: The REF submission provides a comprehensive analysis across various
dimensions, including economic, health, and policy impacts, supported by specific examples and
quantitative data. The AI-generated impacts cover similar themes but are less specific,
particularly in relation to financial and time-based efficiency gains. Additionally, the REF's

detailed organizational case studies and individual stakeholder feedback are missing from the AI-generated version, which reduces the comprehensiveness of the AI impacts. Expanding on specific case examples and quantitative elements could bridge this gap.

"
"13. Study Title : Ensuring the Fair Treatment of Open Banking Customers
Rater 1 Rating = 3: Good
CR Justification: The AI-generated report broadly covered significant aspects of the research impact, including social, economic, and policy dimensions of Open Banking. However, while it touched on the major themes, it did not delve into as much detail as the REF submission on the specific regulatory changes and their ongoing implementation.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated content covered many significant aspects, including regulatory changes and dynamic consent in Open Banking, but lacked depth in specific examples of FCA engagement and details of the consumer forum insights provided in the REF version.
Rater 3 Rating = 4: Very Good
CR Justification: Both versions cover multiple dimensions of impact, including policy, social, and technological implications. However, the REF version provided additional detail on the impacts at specific stages, such as early consultations and gradual regulatory shifts. The AI version could have benefited from more granular detail on these steps, though it is thorough in addressing main points.
Rater 4 Rating = 3
CR Justification: The AI-generated version covers a wide range of impact domains, including social, technological, and regulatory. The REF submission, however, provides more detail in each dimension, particularly in terms of exact policy adjustments and specific regulatory changes. For instance, the REF includes finer details about consumer feedback and responses from regulatory bodies over time, which add depth.

"
"14. Study Title : Advancing Movement Practices in Doctoral and Professional Contexts
Rater 1 Rating = 4: Very Good
CR Justification: The AI-generated report captures a wide range of the key impact dimensions—academic, social, economic, and cultural—mentioned in the REF submission. However, the comprehensiveness of stakeholder engagement and international case studies is richer in the REF version, especially regarding collaborations with Dance4 and other institutions.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated impact is comprehensive in addressing various dimensions of impact, including social, academic, and cultural aspects. It acknowledges the influence on creative practices and educational frameworks. However, it does not fully capture the variety of documented evidence provided in the REF submission regarding the geographic and institutional scope of the CAP's application, such as the detailed examples of specific artists or PhD candidates involved.
Rater 3 Rating = 3 (Good)
CR Justification: While the AI-generated content covers many of the significant impact areas, it omits some details on specific training events, user statistics, and feedback metrics, which are present in the REF-submitted version. The AI version addresses social, educational, and artistic domains but lacks the comprehensive stakeholder engagement insights.
Rater 4 Rating = 3 (Good)
CR Justification: While the AI-generated impacts cover major domains, it does not capture the full scope of CAP's reach within arts and HEI. The REF submission details CAP's integration into artistic doctoral networks and training within European HEIs, including the ADiE network, which extends CAP's impact in academia and policy advocacy.

"
"15. Study Title : Digital Twin Specification, Design and Application
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers many significant dimensions of the impact, including technological, economic, and academic impacts. However, some areas, such as the specific industry examples (e.g., COVID-19 simulations), and collaborations were not fully explored.
Rater 2 Rating = 3: Good
CR Justification: While the AI-generated content captures significant aspects of the research, such as industry collaborations, simulation technology, and structural health monitoring, it misses specific project details. For instance, the REF submission includes detailed use cases such as optimization in retail supply chains and pandemic modeling.
Rater 3 Rating = 4

CR Justification: The AI-generated impacts broadly cover significant aspects of DT impacts, including enterprise modeling, collaboration, and the development of specific products like TwinX™. It captures technological, social, and policy impacts effectively, though some nuanced contextual details, particularly around stakeholder engagement processes and region-specific implementations, are limited. The AI content also attempts to categorize impacts within multiple fields but could expand on specific measurable impacts.
Rater 4 Rating = 4 (Very Good)
CR Justification: The ChatGPT impacts capture the major domains identified in the REF submission, including the technological, economic, and social impacts of digital twin research. The AI version presents these areas in a way that is broadly representative of the REF document. However, certain contextual details, such as the structural health monitoring applications and their contributions to Vietnamese infrastructure, are less elaborately covered.

"
"16. Study Title : Being in Touch: Inspiring Cultural Engagement through Creative-Critical Writing
Rater 1 Rating = 4: Very Good
CR Justification: The AI-generated version covers most key dimensions, including academic, cultural, and social impacts. It successfully identifies the broader interdisciplinary relevance, particularly in educational and sensory studies. However, the level of detail regarding stakeholder engagement and specific workshops/events from the REF submission is slightly lacking.
Rater 2 Rating = 4 (Very Good)
CR Justification: The ChatGPT-generated content covers many of the dimensions of impact in the original REF submission, including academic, cultural, and technological impacts. However, the REF version provides a more comprehensive exploration of specific events, workshops, and the influence on diverse communities, which was somewhat condensed in the AI version.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI-generated content generally covers significant impact aspects, including artistic and cultural contributions, institutional partnerships, and international community engagement. It does an effective job of capturing the project's broad impact, including community and cultural programming. However, some specific community engagement details, such as workshop topics and audience demographics, are less explicit in the AI version.
Rater 4 Rating = 3: Good
CR Justification: The AI-generated content covers a substantial portion of the study's social, cultural, and educational impact areas, adequately reflecting the thematic and methodological diversity highlighted in the REF submission. However, it lacks specific coverage of demographic diversity among participants and does not address Jackson's iterative workshop designs and distinct engagement approaches (e.g., collaboration with UNESCO Creative Cities, the specific structure of workshops). The REF submission comprehensively highlights engagement efforts, such as diversity in demographic participation and thematic workshop structures across locations, providing a fuller context for assessing impact depth.

"
"17. Study Title : Shaping crime prevention policy and strategy to sustain the crime drop and reduce domestic burglary
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers most of the major dimensions of research impact, including policy, social, and academic domains. However, the comprehensiveness is slightly lower because certain specific local and regional initiatives, such as the Nottingham Crime and Drug Partnership, are less detailed in the AI version. The AI version offers a broader but less granular view of the real-world applications mentioned in the REF submission.
Rater 2 Rating = 4 (Very Good)
CR Justification: Both versions cover a wide range of relevant impact areas, including social, policy, and economic dimensions. The AI-generated version addresses critical domains such as public safety and law enforcement, while also reflecting on economic impacts through crime prevention cost savings. However, certain localized impacts, such as specific contributions to regional initiatives like the Nottingham Crime and Drugs Partnership, were more comprehensively detailed in the REF submission.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated content covers all significant aspects mentioned in the original case studies, providing a detailed exploration of both the short-term and long-term impacts, direct and indirect stakeholders, and across multiple domains like social, economic, and policy.
Rater 4 Rating = [4] Very Good
CR Justification: ChatGPT's generated content covers significant dimensions of impact (social,

economic, policy) found in the original submission. While it lacks some specific policy interactions, the coverage is substantial and addresses multiple key areas of impact outlined in the original case study, especially around the social and economic consequences of reduced crime rates.

"

"18. Study Title : Building local socio-economic impacts into the assessment of major energy projects

Rater 1 Rating = 4: Very Good

CR Justification: The ChatGPT version covered a wide range of research impacts, addressing social, economic, policy, and environmental domains. The REF submission, however, had more focused examples, such as the immediate changes made to employment and housing due to the construction of HPC and SZB. Additionally, the REF document went into greater depth on specific socio-economic impacts, such as the use of Community Benefits Agreements and real-time crime management during the construction phases, which ChatGPT lacked.

Rater 2 Rating = 3 (Good)

CR Justification: The AI-generated content covers significant aspects such as local employment, housing, and community impacts. However, the REF submission provides a more detailed and granular analysis of stakeholder feedback, including examples from EDF, government agencies, and local authorities.

Rater 3 Rating = 4: Very Good

CR Justification: The AI content extensively covers the socio-economic, policy, and academic impacts, mirroring the comprehensive nature of the original study but with added generalizations that broaden the impact scope.

Rater 4 Rating = 5 (Excellent)

CR Justification: ChatGPT's impacts thoroughly cover the various socio-economic and environmental dimensions detailed in the REF, including economic stimulation, housing impacts, and local employment. Both versions encompass similar dimensions of impact (e.g., policy, academic, and community impacts), with the ChatGPT version providing expanded insights into sustainability and global transferability beyond local project details.

"

"19. Study Title : Strengthening global and national policies on performance-based and innovative health financing in low-income and fragile settings

Rater 1 Rating = 3 - Good

CR Justification: The AI-generated impacts cover significant dimensions of impact, including social, economic, and health domains. However, the REF submission goes into greater depth on some specific impacts, particularly in fragile settings like Zimbabwe and Sierra Leone, and discusses the impact of the Ebola epidemic on health systems. ChatGPT focuses more broadly on the scalability and potential applications of PBF models but lacks some of the fine detail regarding specific country-level policy outcomes.

Rater 2 Rating = 3 (Good)

CR Justification: The AI version covers a significant range of impacts, from global policy influence to practical implementation. However, it lacks depth in discussing the nuanced social, economic, and technological impacts that were thoroughly discussed in the REF submission. For example, specific donor collaborations and multi-country case studies (e.g., Zimbabwe and Sierra Leone) were more elaborately presented in the REF document.

Rater 3 Rating = 4 (Very Good)

CR Justification: ChatGPT's analysis comprehensively covers significant impact dimensions, including short-term applications in healthcare delivery, long-term foundational impact, and educational impacts within LMIC health sectors. The REF submission further elaborates on specific collaborative impacts with international organizations and on-the-ground adaptation details, which gives it a slight edge in scope. Both versions thoroughly address the policy and healthcare domains but with differing detail levels on specific stakeholder engagements.

Rater 4 Rating = 4 - Very Good

CR Justification: The AI-generated impacts broadly cover the dimensions of social, economic, and technological impact as outlined in the REF study, addressing policy changes, funding priorities, and organizational influence. While the AI report identifies the main areas of stakeholder impact, it lacks certain in-depth references (e.g., specific meeting details and training outcomes) available in the REF submission, particularly regarding WHO's training sessions and the World Bank's discussion forums.

"

"20. Study Title : Transforming the Accessibility and Discoverability of Millions of Archival Television Programmes

Rater 1 Rating = 3 (Good)

CR Justification: The AI-generated content covers many key aspects, including the academic, cultural, and policy impacts, along with the technological innovations associated with TV archives. However, the REF version includes more specific and practical examples, such as collaborative projects and the direct benefits to specific educational and archival stakeholders, which are somewhat generalized in the ChatGPT output.
Rater 2 Rating = 4: Very Good
CR Justification: The AI-generated content provides a comprehensive view of the project's impact across social, academic, and technological domains. However, it sometimes misses the depth of explanation seen in the REF submission about the specifics of user feedback and the precise evolution of the BoB platform. Nonetheless, the AI version extends the potential impacts by considering long-term future applications.
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated impacts adequately cover the major dimensions, like the economic, social, and educational domains. However, the REF submission provides more detail on practical implementations, such as hands-on history methods and user engagement for ongoing platform development. AI content partially addresses scalability but lacks emphasis on the evolving academic contributions within different broadcasting sectors, which the REF highlights extensively.
Rater 4 Rating = 4: Very Good
CR Justification: The AI-generated impacts cover nearly all major impact domains discussed in the REF 2021 submission, including academic, cultural, and technological domains. However, some details regarding specific stakeholders, like archivists and the diversity of educational institutions benefiting from the archives, were more elaborated in the REF version. Overall, the AI report's breadth is high, aligning with the social and cultural implications comprehensively but with less depth on technical metadata innovations.

"
"21. Study Title : Improving quality of life for patients with Parkinson's disease
Rater 1 Rating = 3 (Good)
CR Justification: The ChatGPT version covers most significant dimensions of the REF submission, including clinical guidelines, economic impacts, and policy changes. However, it does not fully elaborate on some stakeholder engagement and detailed clinical trial information presented in the REF version.
Rater 2 Rating = 4
CR Justification: Covers all major dimensions but lacks depth in cost-saving analysis and patient outcome statistics.
Rater 3 Rating = 3 (Good)
CR Justification: The AI version covers the major impact dimensions, including health, economic, social, and policy implications, closely mirroring the REF submission's emphasis on enhanced treatment options and cost reduction. However, the REF version provides more nuanced discussions about the longitudinal study impact and detailed information on surgery outcomes. The AI version generalizes these aspects, offering less focus on surgical interventions and detailed cost-effectiveness findings specifically tied to the NICE and Canadian guidelines.
Rater 4 Rating = 4
CR Justification: The AI-generated content comprehensively addresses the impact areas in clinical practice, policy, and economic dimensions, closely reflecting the depth and breadth noted in the original study. Coverage in both versions spans the healthcare impact domain, especially clinical improvements in PD treatment. However, the REF submission provides exhaustive evidence linking each impact to specific trials and studies, adding a layer of empirical detail that strengthens comprehensiveness.

"
"22. Study Title : Improving Treatment for Women Suffering from Endometrial Hyperplasia
Rater 1 Rating = 3 (Good)
CR Justification: The AI version covers many important aspects of the research impact, including health benefits, clinical guidelines, and sustainability. However, it misses out on the granularity of statistics and specific geographical mentions (like Hong Kong) present in the REF version, which gives it a slightly less comprehensive view of the overall impact.
Rater 2 Rating = [Rating: 3]
CR Justification: While the AI-generated version covers key aspects of the research impact, such as healthcare improvements and policy changes, it does not fully address all dimensions outlined in the REF submission, such as the environmental impact of reduced surgeries (resource savings) or specific NHS cost savings. The comprehensive scope of the REF version, which includes both national and international guideline changes, patient well-being, and long-term health trends, is more detailed.
Rater 3 Rating = 4

CR Justification: The ChatGPT-generated impacts cover most relevant dimensions, including social, economic, and technological aspects, aligning with REF's focus on clinical guidelines, patient outcomes, and cost savings. It occasionally expands upon REF data, offering implications for other healthcare settings. While comprehensive, it lacks some specific details, such as updates in practitioner guidelines globally and exact adoption figures in clinical practice, limiting the depth in some areas.
Rater 4 Rating = 4 (Very Good)
CR Justification: The AI-generated version adequately covers the major dimensions of impact as presented in the REF version, addressing health and economic benefits of alternative EH treatments and their broader application in clinical practice. It also includes broader themes relevant to clinical guidelines but lacks some intricate elements such as shifts in patient behavioral responses (e.g., willingness to adopt non-surgical options) and a comprehensive outline of the detailed impact on practitioners' adherence to new guidelines.

"
"23. Study Title : Heritage of the first farmers
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated impacts capture most significant aspects of the original submission, such as heritage preservation, educational outreach, and museum collaboration. However, some finer details, particularly specific stakeholder feedback and certain capacity-building efforts, are less thoroughly addressed.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated content covers many dimensions of the research impact, including academic, cultural, environmental, and educational impacts. However, while the AI-generated report discusses broader aspects such as potential future trends in sustainable practices, it does not fully capture some of the localized social impacts and specific educational contributions described in the REF submission, such as the detailed interaction with Turkish primary schools.
Rater 3 Rating = 3 (Good)
CR Justification: ChatGPT's version includes most major dimensions—social, educational, economic, and cultural impacts—covered in the REF submission. However, the REF version provides a greater depth in specific areas, such as community-specific feedback and cultural identity preservation, as well as capacity-building initiatives like staff training at the site. These details create a more granular account of project reach and local influence, making the REF version slightly more comprehensive.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated version covers key aspects of the project's research impact, such as cultural, educational, and academic impacts. It captures the overarching scope of archaeological, social, and preservation impacts but lacks the in-depth specificity provided in the REF report about training programs for museum staff and visitor engagement feedback. Furthermore, AI coverage lacks the explicit mention of capacity-building activities and collaborations with Turkish archaeological authorities, which are crucial dimensions noted in the original submission.

"
"24. Study Title : Mathematical modelling of an aneurysm sealing system triggers patient safety policy that withdraws surgical practice from the NHS
Rater 1 Rating = 3
CR Justification: The AI-generated version touches on the major aspects of the research, including patient safety, regulatory impacts, and health benefits. However, it lacks the depth of detail found in the REF submission, particularly around specific clinical data and step-by-step explanation of patient surveillance changes. The REF version is more thorough in presenting regulatory changes and patient numbers impacted.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated content covers the essential dimensions of the impact: health, regulatory, and patient safety. However, it lacks some specific numerical data and patient outcome statistics that were provided in the REF submission. For example, the REF submission includes exact figures for the number of patients affected and statistical comparisons between different procedures, which enhance the comprehensiveness.
Rater 3 Rating = 4
CR Justification: ChatGPT's content covers the primary domains of impact, including regulatory policy, healthcare improvements, and public safety benefits. However, the REF 2021 submission includes additional context on the development of the EVAS system and patient monitoring specifics, which deepens understanding but is absent in ChatGPT's version. ChatGPT generally captures all critical aspects but lacks the granular details in REF's detailed patient outcome statistics.

Rater 4 Rating = 3 - Good
CR Justification: The AI-generated version covers the primary dimensions of impact, including health, policy, and patient outcomes. However, it lacks depth in detailing the direct clinical application aspects such as surveillance protocols and the broader clinical ramifications in specific NHS centers. The REF submission includes more details on clinician testimonials and precise policy mechanisms (e.g., specific NHS guidance).

"
"25. Study Title : Financial and efficiency improvements from socio-technical digitalization of costing and procurement in the built environment
Rater 1 Rating = 4
CR Justification: The AI-generated report comprehensively covers the major themes of digitalization, construction industry improvements, and stakeholder involvement. However, while the technological and economic dimensions are well addressed, the AI version does not cover all significant impacts, such as detailed policy influence and specific stakeholder benefits that are emphasized in the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated version covers many significant aspects of the research, including financial, social, and technological impacts. However, it lacks detail in certain economic outcomes, like the exact monetary savings highlighted in the REF submission.
Rater 3 Rating = 3: Good
CR Justification: The AI-generated version provides substantial coverage of the research's economic, technological, and social dimensions. It successfully touches on industry impact, scalability, and stakeholder relevance, emphasizing practical applications within the construction industry. However, some significant areas, such as the role in influencing government procurement policies and depth in financial implications, are more thoroughly covered in the REF submission, which details specific impacts like cost efficiencies and contract performance improvements.
Rater 4 Rating = 4 - Very Good
CR Justification: The AI-generated report addresses significant domains of impact, including economic, technological, and policy impacts. It covers Boyd's contribution to digital transformation across the supply chain, the influence on procurement practices, and broad policy changes, though it lacks some finer details present in the REF version. For example, the REF submission specifies interactions with high-level stakeholders and provides granular cost savings. The AI version could enhance its comprehensiveness by explicitly connecting specific stakeholders' testimonials to the achieved outcomes.

"
"26. Study Title : Evaluating Effectiveness
Rater 1 Rating = 4 - Very Good
CR Justification: The AI-generated content covers most of the significant aspects of impact mentioned in the REF submission. It highlights domains such as health, education, and policy, with potential long-term impacts. However, the AI version is less detailed in specific institutional collaborations (e.g., CEDIL, World Bank) and particular case studies used to support claims in the REF submission.
Rater 2 Rating = 3 - Good
CR Justification: The AI-generated version covered the major areas of research impact, such as international development, healthcare, and education, similar to the REF submission. However, the REF submission was more comprehensive in discussing specific case studies, such as Cartwright's direct influence on UK educational policy and healthcare guidelines, particularly in terms of her collaboration with specific organizations like CEDIL and NICE.
Rater 3 Rating = 5 (Excellent)
CR Justification: ChatGPT's analysis covered all significant dimensions of impact as in the REF-submitted content, including domains such as social, economic, and technological, with a strong focus on the academic and healthcare domains. It provided a comprehensive account that included potential applications and broader implications across various stakeholders and sectors.
Rater 4 Rating = 4 (Very Good)
CR Justification: ChatGPT captures significant dimensions such as the application of middle-range theory across healthcare, policy, and education. The AI version includes all major sectors impacted by Cartwright's research, though the REF submission includes more granular details on the economic implications of Cartwright's policy critiques, such as the direct influence on specific policy documents and evaluations.

"
"27. Study Title : Seeing beyond the wheelchair: Pioneering education and higher aspiration promotion for boys and men with Duchenne Muscular Dystrophy

Rater 1 Rating = 3 (Good)
CR Justification: ChatGPT covers significant areas but lacks depth in detailing specific policy frameworks and stakeholder collaborations.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated version captures the major domains of impact across health, education, and social services, reflecting the comprehensive coverage of the REF submission. However, it falls slightly short in providing the detailed narrative around stakeholder engagement and direct participation outcomes (e.g., attendance figures at conferences, details of educational materials distributed), which are more thoroughly covered in the REF submission. Additionally, the AI version missed some minor but impactful contributions, such as specific legal advocacy for families of DMD children, which adds to the completeness in the REF case study.
Rater 3 Rating = 5
CR Justification: The AI version covers most aspects of impact thoroughly, including domains of social and health dimensions, intervention types, and projected long-term societal impacts. This demonstrates a comprehensive understanding of the research's goals and its cross-disciplinary influences on education, healthcare, and social inclusion. The ChatGPT content is comparable in depth to the REF version and addresses future implications of the research adequately.
Rater 4 Rating = 3 (Good)
CR Justification: ChatGPT's version covers major domains such as health, social, educational, and policy impacts. However, the REF submission includes deeper information on policy influence, direct community engagement metrics, and data points showing stakeholder reach, which enhance the breadth of the REF's impact case study.

"

"28. Study Title : The Coffee Historian: Achieving Impact Through Industry Collaboration, Education, and Public Engagement
Rater 1 Rating = 4 (Very Good)
CR Justification: The ChatGPT version addresses multiple dimensions of impact, including economic, cultural, and technological aspects. It captures the essence of the global spread of Italian coffee culture and innovations in espresso machines. However, the REF version goes into more detail about specific public engagements and educational outputs, making it slightly more comprehensive in certain areas.
Rater 2 Rating = 4 - Very Good
CR Justification: ChatGPT covers most key dimensions of impact mentioned in the REF case, such as social, economic, and cultural domains, along with technological innovations (espresso machine development) and educational influence within the coffee sector. However, some nuanced elements—like the specific long-term impacts of Morris's public engagements—are less comprehensively developed in the ChatGPT version.
Rater 3 Rating = 4 (Very Good)
CR Justification: The ChatGPT version covers almost all impact areas noted in the REF, including the social, economic, and cultural impacts. However, it lacks some depth in describing individual industry engagements and the broader historical context provided in the REF version.
Rater 4 Rating = 5 (Excellent)
CR Justification: The AI-generated impacts provide a comprehensive view that aligns well with the case study's depth across cultural, economic, and educational impacts. The ChatGPT version touches on most significant aspects discussed in the REF submission, including detailed coverage of educational outreach and specific industry collaborations. The AI report extends its coverage by emphasizing long-term industry impacts and scalability of coffee culture globally, which further adds to the thoroughness.

"

"29. Study Title : Discovering Ted Hughes's Yorkshire
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI version covers a wide range of potential impacts, including cultural, educational, and even psychological dimensions of Hughes's work. It explores the academic and societal implications thoroughly. However, some detailed collaborations, the scope of practical local impacts (like trail maps, cultural tourism, and public activities), and specific funding mechanisms were better covered in the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: The ChatGPT version covers many significant impact areas mentioned in the REF submission, such as cultural tourism, local economic development, and community engagement through the THN. However, it does not capture the depth of stakeholder interactions or the specific long-term initiatives (e.g., the heritage trails and educational programs) as thoroughly as the REF submission.
Rater 3 Rating = 4

CR Justification: The AI-generated content comprehensively covers most aspects of the REF submission, including community impact, cultural participation, and the development of literary trails, aligning with the overall objectives of enhancing Ted Hughes's legacy in Yorkshire.
Rater 4 Rating = 4
CR Justification: The ChatGPT-generated content broadly covers the multiple impact dimensions—cultural, social, and economic—as presented in the REF submission. However, there is limited mention of the academic advancements and internal network development that THN fostered, which were emphasized in the REF.

"
"30. Study Title : Empowering Indigenous Self-Representation for the Emberá People of Panama
Rater 1 Rating = 3 (Good)
CR Justification: The REF submission is more comprehensive, detailing not only the impacts on tourism but also the political, legal, and socio-economic dimensions of the research. The AI-generated version covers many of these aspects but lacks in-depth examples, particularly regarding political engagement and specific stakeholders (e.g., NGOs and policymakers).
Rater 2 Rating = 3: Good
CR Justification: The ChatGPT-generated content captures most dimensions of the research impact, including social, economic, and cultural domains. However, it was less comprehensive in detailing the procedural legal steps and the specific timeline of impacts related to land claims and modern cultural practices mentioned in the REF submission.
Rater 3 Rating = 3: Good
CR Justification: The AI-generated content sufficiently covers the main impacts on cultural, economic, and social fronts, aligning with the broader impacts detailed in the REF submission. However, it lacks depth in the incremental development aspects of the tourism narrative, such as how these narratives evolved over time to reflect Indigenous modernity alongside traditional elements. Furthermore, the sequential involvement of community leaders and younger generations in redefining cultural authenticity, as emphasized in the REF, is not as thoroughly developed in the AI version.
Rater 4 Rating = [5] Excellent
CR Justification: ChatGPT version covers significant aspects, including social, economic, and cultural impacts, while additionally providing a technological aspect on tools for self-representation. This thoroughness mirrors REF's multidimensional impact narrative, demonstrating an exhaustive approach to the project's breadth.

"
"31. Study Title : Enhancing Understanding of the Foreign and Security Policy Implications of Brexit for Government and the Wider Policy Community
Rater 1 Rating = 3 (Good)
CR Justification: The REF submission offers a detailed account of various stakeholder groups and how Whitman's research directly influenced Brexit-related discussions, policy, and public understanding. ChatGPT captures the broader impacts but lacks some of the finer details, especially in terms of specific stakeholder engagements and detailed instances of media outreach.
Rater 2 Rating = 4 - Very Good
CR Justification: The AI-generated report covers a comprehensive range of impact domains—academic, policy, social, and economic. It aligns with the multidimensional impact noted in the REF submission, including effects on UK-EU relations, policy development, and public discourse. However, it does not fully match the level of detailed stakeholder engagement found in the REF submission, especially specific examples of government and media involvement.
Rater 3 Rating = 3
CR Justification: The AI-generated report covers most of the REF submission's key impacts, addressing dimensions like public understanding, policy influence, and collaborations. However, it lacks comprehensive detail in terms of specific stakeholders and Whitman's engagement with foreign policy decision-makers in both the UK and EU. While it presents a good overview, it omits certain specifics found in the REF version, such as the international briefings in Australia and New Zealand.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated impacts capture major impact areas, including policy, social, and economic dimensions. However, it falls short in detailing specific stakeholder interactions and long-term follow-up measures outlined in the REF submission, which include targeted international briefings and bilateral consultations.

"
"32. Study Title : Worldwide Improvements in Policing due to Increased Sales of Facial Composite Software
Rater 1 Rating = 3 (Good)

CR Justification: The ChatGPT version covers the primary dimensions of impact (technological, academic, social, and economic), but it lacks the level of comprehensiveness seen in the REF submission, especially regarding the depth of user testimonials, precise international deployment numbers, and training programs. While the AI-generated content highlights the impact on law enforcement, it misses specific details about stakeholder engagement and international training programs.
Rater 2 Rating = 3
CR Justification: While the AI-generated version addresses multiple dimensions such as social, technological, and academic impact, it lacks the depth found in the REF submission, especially in the economic and social-cultural areas. The AI version broadly outlines these impacts but misses specific outcomes like detailed financial growth or media appearances of the software.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI-generated impacts cover most key impact dimensions—such as social, economic, and technological areas—similar to the REF submission. Additionally, they capture the system's scalability and adaptability, though without the specific case-based examples. AI content addresses a variety of impacts but lacks certain detailed examples from real-world adoption cases seen in the REF version.
Rater 4 Rating = 5: Excellent
CR Justification: The AI version covers the main aspects described in the REF submission, including international adoption, commercial success, training impacts, and cultural relevance. All significant impact dimensions—social, economic, technological—are comprehensively addressed.

"
"33. Study Title : Improving the Physical Wellbeing of the Police Force
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI-generated report covers most of the significant dimensions like social, health, and policy impacts. The scope of impact is well-represented, although certain program-specific evaluations or interim data findings could have been more extensively covered.
Rater 2 Rating = 4: Very Good
CR Justification: Both versions provide comprehensive coverage of the physical well-being initiative. The REF submission goes into greater depth regarding program logistics, interim evaluations, and the specific impacts on police staff. The AI version addresses the key components but lacks some details, such as post-program health outcomes and the program's expansion to other forces.
Rater 3 Rating = 3
CR Justification: The AI-generated content covers key dimensions such as health, policy, and workplace wellbeing but does not fully capture the in-depth development process and multi-level stakeholder engagement mentioned in the REF submission. Specifics of fitness mentor training, handbook development, and post-program support are less prominent.
Rater 4 Rating = 3 (Good)
CR Justification: The AI content covers the major areas of impact, focusing on physical health improvements and police well-being. It captures policy, community, and health dimensions, reflecting the main aspects from the REF document. However, there is a noticeable gap in capturing some of the operational specifics and post-program feedback extensively discussed in the REF submission, affecting overall comprehensiveness.

"
"34. Study Title : Raising Maori students' achievement in secondary schools in Aotearoa New Zealand
Rater 1 Rating = 3
CR Justification: While the ChatGPT-generated content captures the core educational and pedagogical impacts, it does not fully explore the scope of societal and policy impacts detailed in the REF submission, particularly the long-term reforms in school leadership and policy integration regarding Māori communities. The AI version also tends to generalize in places where the REF submission provides specific examples of quantitative outcomes.
Rater 2 Rating = 3 (Good)
CR Justification: The REF submission provides a more comprehensive view of the project's impacts, including specific statistics on student achievement, teacher training, and community engagement. The AI-generated content captures most of the significant impacts, but lacks the depth and specificity regarding the implementation and measured outcomes in schools. For example, the REF submission goes into detail about the number of schools and teachers impacted by the KEP initiative, which is not present in the AI-generated version.
Rater 3 Rating = 4 (Very Good)
CR Justification: The ChatGPT analysis addressed almost all significant dimensions, like social, economic, and educational impacts. However, some specific quantitative data related to teacher and student participation were missing, reducing comprehensiveness slightly.

Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated content covered the main dimensions: social, educational, and community relationships. However, while the breadth of impacts was represented, specific data points (e.g., 70% of schools reporting ownership in transformative changes) were missing or only generally referenced, reducing the coverage compared to the more detailed quantitative data found in the REF submission.

"

"35. Study Title : Improving patient outcomes and treatment guidelines through the study of Hepatitis C
Rater 1 Rating = 3 (Good)
CR Justification: The ChatGPT version covers significant aspects like healthcare improvements and policy influence but lacks depth in areas such as specific national programs (e.g., HepCATT) and detailed patient outcomes, which are extensively described in the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: The AI version generally covers the major impact dimensions, including social, clinical, and policy aspects of the research. However, it does not fully cover every aspect as comprehensively as the REF submission, which provided additional layers of depth on policy influence, specific national outcomes, and engagement with multiple stakeholders. For instance, the AI report captures the role of DAAs and patient outcomes but does not explore certain interventions such as 'reflex testing'.
Rater 3 Rating = 4
CR Justification: The AI report includes most of the REF-detailed impacts, such as improved HCV patient care pathways, policy influence, and international guideline contributions, covering essential dimensions like health outcomes and healthcare system impacts. However, it omits granular information on specific institutions, stakeholders, and quantitative data.
Rater 4 Rating = 3 - Good
CR Justification: The AI-generated content covers many core aspects of impact discussed in the REF submission, including patient outcome improvements, national policy guidance, and cost-effective treatments. However, some key details are missing, such as specific descriptions of national guideline contributions, detailed pharmaceutical partnerships, and targeted patient demographics (e.g., PWID, patients with advanced liver disease). Additionally, AI content does not fully cover the cascade of care framework that plays a prominent role in the REF submission.

"

"36. Study Title : Improving Homecare Quality in the UK Through Optimized Workforce Planning
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated impacts cover most aspects mentioned in the REF submission, including economic and technological benefits and workforce well-being. However, it lacks some of the detailed industry examples provided in the REF submission, particularly real-world case studies like Cheshire West Council or Apex Prime Care.
Rater 2 Rating = 3 (Good)
CR Justification: Comprehensive but misses qualitative details and specific client testimonials during COVID-19.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated content comprehensively addressed the major impact dimensions: economic, societal, and technological, in line with the REF-submitted case study. Additional potential dimensions were noted, suggesting further scalability and relevance across industries.
Rater 4 Rating = 4 (Very Good)
CR Justification: The AI-generated impact statements cover most significant aspects, including economic, technological, and social impacts. The collaboration's commercial success and scalability in workforce management are both highlighted. However, while covering most dimensions, the AI-generated report was slightly less detailed in specific examples related to employee satisfaction and retention metrics.

"

"37. Study Title : Transforming vaccine policy for pneumococcal disease leading to significant cost savings in the NHS
Rater 1 Rating = 3 - Good
CR Justification: The ChatGPT-generated content covers several key dimensions, such as health, technological, and policy impacts. However, the depth of the coverage is somewhat less detailed compared to the REF submission, which provides specific examples of influence on JCVI and ACIP decisions, as well as NHS cost savings and the impact of childhood vaccination policies. ChatGPT mentions general dimensions like public health and diagnostics, but it does not delve into the intricate layers of how the research influenced specific committees or vaccine schedules.

Rater 2 Rating = 4: Very Good
CR Justification: The AI version covers most of the REF submission's major points, including health policy and NHS impacts, but may miss specific details on childhood vaccination and exact economic impacts.
Rater 3 Rating = 4: Very Good
CR Justification: The AI-generated content broadly captures the key domains of health, policy, and economic impact but lacks in-depth details on specific UK policy recommendations. It covered more generalized global impacts, aligning well with the REF submission's objectives.
Rater 4 Rating = 3
CR Justification: The AI version includes key areas like policy impact and economic savings but omits some specifics on the influence of Nottingham's unique longitudinal study and the study's specific advancements in diagnostics. It broadly covers policy influence, economic impact, and health outcomes, yet lacks specific mentions of other aspects, such as U.S. policy changes or ongoing model adjustments referenced in the REF submission.

"
"38. Study Title : UoP32Househistories: A House Through Time: Shaping a flagship TV series to achieve critical and financial success and inspiring the public to engage with house history
Rater 1 Rating = 3 (Good)
CR Justification: The ChatGPT-generated version adequately covers the main research impacts described in the REF submission, including public engagement, academic influence, and social impact. However, it does not fully capture the breadth of collaborations with industry professionals (such as FindMyPast) and the precise contributions to heritage conservation that the REF version details. The ChatGPT version tends to generalize some impacts that were specified in the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated version covers most of the essential dimensions of the research impact, including social, economic, and public engagement impacts. However, it lacks the full range of detailed examples and nuances provided in the REF submission, particularly in terms of the series' ongoing academic and public discourse.
Rater 3 Rating = 3 (Good)
CR Justification: The ChatGPT version covers the main research impacts such as cultural influence, historical preservation, and social engagement. However, the REF submission contains additional details, including specific contributions to the series, further collaborations, and in-depth descriptions of each research component.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated impacts capture many significant elements from the REF submission, such as the influence on public interest in house history and the specific impact on BBC Two. However, the depth of specific stakeholder engagement and historical detail seen in the REF submission is somewhat generalized in the AI output, which could be enhanced with specifics.

"
"39. Study Title : Optimising baggage operations at London Heathrow Airport to achieve cost savings for the aviation industry
Rater 1 Rating = 3 (Good)
CR Justification: The REF submission provides a focused and detailed account of the impacts of the project, with in-depth analysis on how the research influenced specific stakeholders like Heathrow Airport, airlines, and passengers. The ChatGPT version expands on multiple potential dimensions such as environmental, technological, and policy impacts. However, the AI-generated version lacks the specific economic details and real-world case examples that make the REF submission more tangible and comprehensive.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated content covered most of the significant aspects of the original research. It mentioned technological, economic, and operational impacts but missed some finer details, such as specific partnerships (e.g., with Arup) and the deeper technical challenges addressed by the project. The original submission also provided concrete examples of impact, including statistical improvements in operations (e.g., the 20% reduction in short-landed bags).
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated content covered the main dimensions of impact, including economic and social implications. However, it missed specific financial figures, such as capital expenditure savings and contract details, which add to the comprehensive understanding of the REF submission.
Rater 4 Rating = 3 - Good
CR Justification: The AI report covers the majority of key impact areas: economic, operational efficiency, and passenger experience. However, some nuanced dimensions specific to Heathrow's

policies and the partnership's history with Arup were less explored, affecting comprehensiveness.

"
"40. Study Title : POWeR – Cost-effective online support for weight management
Rater 1 Rating = 3
CR Justification: While the ChatGPT-generated report touches on a broad range of dimensions, it lacks some specific details presented in the REF submission, especially around long-term partnerships and contractual dissemination of the POWeR system. The REF version provides a more comprehensive explanation of both pre- and post-publication impact pathways.
Rater 2 Rating = 4: Very Good
CR Justification: The AI version covered most of the critical dimensions of the POWeR project's impact, such as the clinical, social, and economic impacts of weight management through digital interventions. It included comprehensive discussions on the NHS context, the public health collaborations, and the clinical outcomes for primary care patients. However, it lacked some of the more specific details related to the partnership with specific local authorities and the mechanics of ESRC and Innovate UK funding, which were well-elaborated in the REF submission.
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated impacts address health, technological, economic, and policy dimensions, covering the broad impact spectrum. It includes potential impacts on long-term health outcomes and primary care innovation. However, the REF version provides a more detailed narrative on each stakeholder, dissemination phase, and community-based application, which was less emphasized in the AI-generated text.
Rater 4 Rating = 3
CR Justification: The REF submission offers a comprehensive account, detailing the phased development, clinical trials, and specific engagement with local authorities and policy stakeholders. The ChatGPT version includes the intervention's broader potential in cost savings and public health impact but lacks comparable specificity in funding sources and geographic implementations.

"
"41. Study Title : The global impact of Sunderland's football research on scouting, training, and player preparation in elite football
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated report covers most of the significant aspects of the research impact dimensions, addressing training improvements, player preparation, and the broader influence of the research on football clubs globally. However, it lacks some of the depth in terms of specific case studies and citations of key organizations and stakeholders mentioned in the REF submission.
Rater 2 Rating = 4: Very Good
CR Justification: The AI-generated content captures the core elements of the research impact, addressing the research's implications across various dimensions such as social, economic, and technological aspects. It thoroughly discusses player performance metrics and the practical applications of the findings in training and coaching.
Rater 3 Rating = 3: Good
CR Justification: The AI-generated report covers the majority of the primary impact areas presented in the REF 2021 submission, such as social, economic, and health domains. It adequately addresses the training and injury-prevention benefits highlighted in the REF submission. However, it lacks specificity in certain areas, such as exact organizations or teams involved, which are pivotal in the REF submission.
Rater 4 Rating = 5 (Excellent)
CR Justification: The ChatGPT report comprehensively covers various dimensions of the impact, spanning from player performance analytics to institutional changes at clubs. The AI content captures long-term outcomes and incorporates both short-term practical applications and broader societal impacts, reflecting an in-depth understanding of the original study's objectives.

"
"42. Study Title : Adding value to convenience retailing through improved pricing, a new store concept, and capacity building
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI-generated impacts are comprehensive in covering various aspects of research impact—economic, social, and policy-related—similar to the REF submission. However, it does not delve deeply into some nuanced details like the organizational changes or specific revenue models mentioned in the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated version covers a wide range of impact areas, including economic, social, and consumer behavioral impacts. However, the comprehensiveness was slightly

limited, as the AI-generated version did not address all nuanced details such as the store's market recognition or the depth of capacity-building activities mentioned in the REF submission.
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover most significant areas of the REF submission, such as sales increase due to convenience initiatives, consumer satisfaction, and operational efficiency improvements. The AI report also included relevant cross-domain impacts. Nevertheless, some areas—such as detailed consumer demographic targeting and specific revenue uplift amounts mentioned in the REF submission—were less comprehensive in the AI version.
Rater 4 Rating = 3
CR Justification: The AI-generated version covers many impact areas, such as economic benefits, consumer satisfaction, and competitive strategy but does not entirely capture the depth and specifics of each dimension. For instance, while consumer behavior research is highlighted, detailed aspects, like adjustments to pricing management structures, are less prominent.

"
"43. Study Title : Transforming Evidence-Based Practice in Public Health Through Co-Production and Evaluation
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI-generated impacts provided a comprehensive coverage of the research's academic, social, and policy impacts, capturing significant areas such as public health, mental health, and education. It also addressed economic and societal aspects. Some specific details and case examples were less explicitly developed compared to the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated report covers most major areas, including mental health interventions, teenage pregnancy, and suicide prevention. However, the REF submission provides a more thorough exploration of policy outcomes, co-produced evaluations, and their implementation within specific local government programs. The AI version lacks some of these finer details, particularly in how these studies influenced both immediate and long-term changes in policy and practice.
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated version covers significant themes, including social, educational, and health impacts. However, certain detailed aspects, such as the exact impact on each program's outcomes and the breadth of specific stakeholder involvement, are less comprehensively covered.
Rater 4 Rating = 5
CR Justification: The AI report covers all significant aspects identified in the REF, including the diverse public health dimensions—mental health, suicide prevention, teenage parent programs, and youth mental health. It also includes future projections, applicability, and potential impact on policy beyond DCC's local scope. This thoroughness provides a high level of detail and comprehensive alignment with the case study.

"
"44. Study Title : Well-being and Public Policy
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers most dimensions of impact, particularly in the areas of public policy, societal influence, and health-related interventions. However, the ChatGPT version does not go into the same depth concerning specific stakeholders (e.g., civil servants) or provide the detailed list of engagements that the REF submission includes. Additionally, international influence, such as Oswald's work with the European Commission, was not as extensively covered in the AI version.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated impact report broadly covers social, economic, and policy impacts, reflecting many of the key dimensions discussed in the REF submission. This includes academic advancements in happiness economics and their application to government policy. However, the AI version does not fully capture all nuances, such as the impact on international policy (e.g., the European Commission), which is addressed in the REF submission. The AI report also lacks the depth of personal testimonies and stakeholder feedback present in the REF, which showcases more diverse real-world applications of the research.
Rater 3 Rating = 3
CR Justification: The AI-generated impacts provide a comprehensive overview of social, economic, and technological implications, but they lack in-depth coverage of unique, direct feedback from government figures on the research's role in national policy changes (e.g., direct ONS consultations and tailored well-being measures for UK citizens). Key specifics, like well-being measurement adaptations, are more thorough in the REF version.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated report covers many of the major dimensions, including social,

economic, and policy-related impacts, but it lacked comprehensive details on some key long-term projects and advisory roles explicitly noted in the REF submission. It provided adequate but less extensive references to the documented shifts in ONS policy, and missed some specific data or project-based insights.

"

"45. Study Title : Improving the quality of green infrastructure in towns and cities in the UK
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated impacts cover a broad range of relevant impact dimensions (e.g., environmental, policy, social, and economic), capturing key outcomes such as improving climate resilience and influencing urban planning. However, the REF version includes additional details regarding specific stakeholder interactions, project scalability, and certification of planning policies (e.g., 30 local authority policies being certified with Building with Nature). The AI-generated version is more generalized and misses some of these details.
Rater 2 Rating = 3
CR Justification: The ChatGPT version addresses the broad dimensions of the research impact, including environmental and societal effects. However, it does not cover all the finer details present in the REF submission, particularly the specific milestones achieved through the project, the geographic spread of the impact, and the detailed framework developed with stakeholders.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated content covers the significant aspects of the impact areas, providing a well-rounded perspective on social, environmental, and economic impact. It also includes discussions on knowledge dissemination pathways and potential for scalability across regions and sectors.
Rater 4 Rating = 3: Good
CR Justification: While covering significant areas such as environmental, social, and policy impacts, the AI-generated version lacks depth in specific examples of policy adaptation in local councils, as seen in the REF's specific mentions of Newcastle and Cheltenham. Furthermore, the AI version does not fully capture the granular influence of partnerships on shaping the standards for green infrastructure.

"

"46. Study Title : Redressing the state of the stateless: seeking political recognition for Tibet and Kashmir
Rater 1 Rating = 4: Very Good
CR Justification: The AI-generated content provides a comprehensive view of the various dimensions of research impact, covering legal, policy, social, and cultural aspects. Both versions cover the asylum rights for Tibetans and the U.S. pressure on India regarding Kashmir. However, the AI version also highlights potential global cultural and policy impacts in ways that the REF submission did not emphasize. It adequately captures the breadth of the impact but misses some deeper contextual nuances regarding the activism of the researchers.
Rater 2 Rating = 4 - Very Good
CR Justification: The AI-generated report covers most significant aspects of the research impact, including legal, political, and academic dimensions. However, it lacks certain specific legal case studies and personal testimonies that are emphasized in the REF submission.
Rater 3 Rating = 3
CR Justification: The AI-generated content covers key impact dimensions such as legal, policy, and social influence but lacks specific examples seen in the REF submission. It includes a broad analysis of advocacy and legal impacts, but certain intricate legal precedents and court interactions that strongly underpin the REF impacts are generalized.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated version provides coverage across social, legal, and policy impact dimensions but lacks the granularity of specific events and figures critical to understanding the depth of influence achieved, as seen in the REF submission. This results in a ""Good"" rating.

"

"47. Study Title : Evidence-based enteral feeding practices for very preterm or very low birth weight infants
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover a range of dimensions including health, social, economic, and technological impacts, but it misses some specific, detailed examples of policy influence that the REF submission provides. The AI report tends to focus more broadly on potential long-term applications.
Rater 2 Rating = 3 (Good)

CR Justification: The ChatGPT-generated version covers the main dimensions of research impact, including clinical, health, and policy implications. However, it does not provide the same level of detail about educational impacts, particularly in training healthcare professionals. The REF submission gives detailed examples of how the research influenced the training and guidelines for healthcare workers, which are underrepresented in the AI-generated version.
Rater 3 Rating = 5: Excellent
CR Justification: The ChatGPT version effectively covers all impact areas, mentioning health improvements, guidelines, cost savings, and global policy influence. It also touches on areas like economic and educational impacts, aligned with the REF details and expanded further, showcasing a holistic impact across domains.
Rater 4 Rating = 3
CR Justification: While the AI-generated impacts broadly cover areas like policy, clinical practice, and patient outcomes, they lack the depth of examples provided in the REF 2021 submission. For instance, the original report highlights explicit increases in milk bank establishments and the integration of recommendations across multiple international guidelines, while the AI version groups impacts more generally.

"
"48. Study Title : Improving Environmental Conservation in East Africa and Beyond
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI-generated content is comprehensive and addresses most dimensions of impact outlined in the REF submission, including environmental, academic, and societal aspects. However, it could expand more on economic impacts, such as eco-tourism development and financial impacts on local communities, which are given considerable attention in the REF version.
Rater 2 Rating = 3
CR Justification: The ChatGPT-generated report covers many significant aspects of the original research but lacks some depth in key areas such as community engagement and the measurable local impact on livelihoods. The original REF submission provides richer detail on the socio-economic benefits, while the AI version more frequently emphasizes technological and methodological advancements.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated report covers all major dimensions of impact, including environmental, economic, policy, and social, as noted in the REF submission. It also addresses both short- and long-term impacts effectively, providing a well-rounded view of the study's impact across stakeholders and timeframes.
Rater 4 Rating = 3: Good
CR Justification: The ChatGPT impact analysis comprehensively covers most conservation domains, from ecosystem management to community impact, albeit with limited depth in specific areas such as revenue generation for local communities. Additionally, while it emphasizes environmental and policy aspects, economic and educational impacts lack detailed analysis, which is more comprehensive in the REF submission. Overall, the coverage of conservation outcomes is adequate but less detailed.

"
"49. Study Title : Predicting the properties of materials with first-principles electronic structure software (CASTEP)
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers the primary dimensions (technological, academic, economic), but it lacks some of the in-depth examples of cross-disciplinary, policy, and specific stakeholder engagement presented in the REF submission. There is also less emphasis on CASTEP's direct economic contributions and real-world use cases, making the coverage slightly less comprehensive.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated content provides a strong overview of the industrial, technological, and economic impacts of CASTEP. However, it lacks some of the specific examples and thorough explanations found in the REF submission, such as the EU-funded study on the return on investment (ROI) for industrial research using CASTEP, and the in-depth case study examples regarding patents and specific industries.
Rater 3 Rating = 3
CR Justification: AI's coverage is comprehensive for major academic and industrial impacts, with generalized insights across domains. However, the AI lacks the detailed specificity in examples, such as ROI figures and exact industries impacted, reducing the report's completeness in fully representing the original REF submission.
Rater 4 Rating = 4 - Very Good
CR Justification: The AI-generated version covered most significant impact areas (academic, technological, and economic) well. It addressed CASTEP's usability, impact on patents, and

collaborations with industries effectively. However, the AI version generalized some impact areas without detailing individual industry use cases or highlighting specific training programs and their reach.

"
"50. Study Title : Statistical pattern recognition applied to protein crystallisation images in the pharmaceutical industry
Rater 1 Rating = 3 - Good
CR Justification: While the ChatGPT version covers a wide range of impacts (academic, technological, health, and economic), it lacks some details about specific implementations and stakeholder collaborations found in the REF 2021 version. The AI-generated version addressed domains like sustainability and scalability, which were only briefly mentioned in the REF submission.
Rater 2 Rating = 3
CR Justification: The AI-generated report covers many key dimensions of impact, such as technological advancements and economic benefits. However, it lacks depth in certain areas that the REF version covers, such as the specific technological processes (e.g., use of the Fourier transform, wavelet analysis) and the integration into specific industrial applications. The AI-generated content also does not fully address the interdisciplinary collaboration discussed in the REF submission.
Rater 3 Rating = 3: Good
CR Justification: The AI-generated version covers major dimensions, including health, economic, technological, and social impact. However, it lacks depth in certain examples and testimonials that were prominent in the REF submission, which highlights specific impacts such as the speed and reliability improvements in high-throughput crystallization. While the ChatGPT version gives a broad overview, it doesn't capture all technical details provided in the REF version, particularly regarding the specific methodological advancements.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated version addresses most of the impact dimensions, covering social, technological, and economic impacts relevant to the project. However, some specific examples of stakeholder endorsements and project applications detailed in the REF submission are summarized or generalized in the AI version.

"
"51. Study Title : 'Moving beyond one-size-fits-all: Improving Widening Participation through Realist Evaluation methodologies in Northern England'
Rater 1 Rating = 3
CR Justification: The AI-generated report covers the significant dimensions of research impact but lacks some of the nuanced depth found in the REF submission. While both versions discuss the role of RE and WP, the REF version is more detailed in describing specific interventions and their outcomes, such as the use of community 'role models.'
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated version covers most major aspects of the research impact: social equity, education, and methodological innovation through RE. It addresses the core goals of widening participation and improving access to higher education. The original REF version, however, is more comprehensive in providing concrete examples of how these impacts were achieved, such as mentoring schemes and community outreach programs tailored to specific regional needs.
Rater 3 Rating = 4
CR Justification: The ChatGPT version covers a comprehensive range of impact dimensions, addressing social, economic, and technological aspects of Widening Participation (WP). However, the REF version provides in-depth local contextual insights on community-specific engagement strategies, which the AI version only generalizes.
Rater 4 Rating = 4 (Very Good)
CR Justification: The AI-generated content provides a comprehensive coverage of the main impact dimensions, such as social equity, institutional collaboration, and community outreach, paralleling the depth of content in the REF report. However, some broader aspects of the WP initiatives, such as the role of ""sense of place"" and detailed evaluations of the mentoring systems, were covered less comprehensively than in the REF submission.

"
"52. Study Title : Challenging monolithic conceptualisations of English for learning, teaching and assessment: The Changing Englishes online course
Rater 1 Rating = 3 - Good
CR Justification: While the AI-generated report covers most of the significant impacts, including social, educational, and policy impacts, it does not delve into some specific reach

and engagement metrics as extensively as the REF submission, which provides data on course usage and regional adoption.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated content covers most major dimensions of the research impact, such as academic, social, and policy influences. It discussed potential changes to ELT practices, language policies, and teacher education. However, it lacked some depth in certain areas, particularly in the social and educational impact dimensions related to concrete examples of course incorporation into global teaching programs as described in the REF submission.
Rater 3 Rating = 3 (Good)
CR Justification: While both versions cover essential aspects such as educational, social, and policy impacts, the REF version provides a more granular breakdown of specific program uptake, geographic reach, and user testimonials, enhancing comprehensiveness. The AI version could improve by expanding on these specifics and including quantitative data on user engagement.
Rater 4 Rating = 4 - Very Good
CR Justification: The AI-generated content covers essential impact areas, including educational reform, policy influence, and social equity in ELT, aligning well with the primary domains of impact mentioned in the REF submission. Some secondary areas, such as individual educator testimonials, were less detailed, which slightly limits the comprehensiveness.

"
"53. Study Title : Robert Paul and the Origins of British Cinema
Rater 1 Rating = 3 (Good)
CR Justification: The ChatGPT version covers most dimensions of the impact, including technological, cultural, and economic effects, but lacks some of the in-depth details on community engagement and direct public awareness efforts seen in the REF submission.
Rater 2 Rating = 4 - Very Good
CR Justification: The AI-generated content covers key research impacts such as technological innovation, preservation of film history, and the cultural impact of early British cinema. However, it lacks some nuanced descriptions of how the research influenced various cultural institutions (e.g., the BFI and the National Science and Media Museum) and how public outreach efforts specifically shaped public and institutional recognition.
Rater 3 Rating = 4 (Very Good)
CR Justification: The ChatGPT version broadly covers technological, economic, cultural, and educational impacts, as seen in the REF 2021 version, addressing the main dimensions. However, it misses some fine details related to specific exhibitions and articles in public media.
Rater 4 Rating = 3: Good
CR Justification: The AI-generated report broadly addresses the cultural, social, and technological impacts noted in the REF submission. However, it does not fully capture the depth of community engagement efforts, such as the impact of specific exhibitions and outreach activities on local venues. Additionally, while it mentions the educational value of the project, it does not detail the tailored educational tools (e.g., comic books) used to reach younger audiences.

"
"54. Study Title : The Care and Management of Gout in Primary Care
Rater 1 Rating = 3 (Good)
CR Justification: While the AI-generated content covers the significant aspects of the research impact, the REF version provides a more comprehensive, detailed account of various stakeholders and their direct engagement with the research outcomes. For instance, it includes detailed descriptions of stakeholder involvement in the development of new guidelines and the role of patient advocacy groups.
Rater 2 Rating = 3: Good
CR Justification: While the AI-generated content covers most major dimensions of impact, such as clinical practices, public health policies, and health outcomes, it does not delve as deeply into all subtopics (e.g., specific public engagement and expert consensus efforts on terminology). The REF submission offers a broader view of patient education efforts, professional training, and the impact of survey reports like Gout Nation. The AI-generated version, though insightful, is not as exhaustive in certain areas.
Rater 3 Rating = 3
CR Justification: The AI-generated report covered substantial areas like patient education, clinical guidelines, and long-term patient outcomes. However, it lacked certain detailed aspects seen in the REF submission, such as regional disparity data and specific guideline implementation methods, which limited its comprehensiveness.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover many key dimensions, such as public policy impact, healthcare improvements, and patient engagement. However, some subtleties, such as

changes in practice beyond gout alone (e.g., broader inflammatory disease management strategies), were less detailed than in the REF version.

"

"55. Study Title : From Victims to Actors: Shifting the Policy Paradigm to Value Children's Contribution in Disaster Risk Management
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated version covered most of the dimensions, such as societal impact, policy influence, and educational reforms, similar to the REF-submitted version. However, certain specific examples and detailed case studies from the REF version were absent. The ChatGPT version was comprehensive in scope but lacked some granularity.
Rater 2 Rating = 3 - Good
CR Justification: The AI-generated report adequately covered major dimensions of impact, including societal, educational, and policy impacts. It addressed the broad scope of child participation, community resilience, and flood preparedness. However, it lacked detailed examples of specific stakeholders and omitted certain educational tools and policy shifts cited in the REF submission. For example, the REF version included extensive documentation on specific interactions with the UK Environment Agency and international stakeholders.
Rater 3 Rating = 4 - Very Good
CR Justification: The AI-generated impact report covers multiple impact dimensions (e.g., social, educational, policy, and community), mirroring the breadth in the REF submission. However, some finer points, like child empowerment in specific public events or detailed stakeholder responses, are more elaborated in the REF document.
Rater 4 Rating = 4
CR Justification: The AI-generated version covers many crucial elements, including the importance of policy, societal awareness, and educational inclusion. However, it lacks some details on methodologies used for engaging children, such as participatory tools, workshops, and partnerships which the REF submission describes comprehensively, particularly in the EU Horizon2020 CUIDAR project context.

"

"56. Study Title : Improved Crab Fisheries Management Benefits Coastal Livelihoods in Brazil
Rater 1 Rating = 3 - Good
CR Justification: Both versions cover important dimensions of the impact, such as economic, social, and policy. However, the REF submission provides more depth on the technical aspects of the prediction methods and their precise application in different regions of Brazil. The REF submission is also more thorough in its discussion of direct engagement with stakeholders like fishers and policymakers. The AI-generated version is solid in discussing the broader social and economic impacts but doesn't cover some nuanced regulatory or technical details.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated report covers most of the significant aspects mentioned in the REF submission, such as environmental, economic, and policy impacts. However, some aspects of stakeholder engagement, like the direct involvement of artisanal fishers and restaurant owners, are not as comprehensively discussed as in the REF submission. The scope of the AI's content does capture the multi-dimensional impact but could have provided more specifics on certain beneficiaries.
Rater 3 Rating = 5 - Excellent
CR Justification: The AI-generated version provides a comprehensive overview, addressing multiple dimensions of impact, including social, economic, environmental, and technological, which are crucial to understanding the research's full scope. The REF submission matches these dimensions, ensuring thoroughness in capturing the various facets of impact.
Rater 4 Rating = 3
CR Justification: The AI-generated content covers the main impact domains, including environmental, economic, and social dimensions. However, the REF version delves into detailed statistics (e.g., population numbers affected, financial impact on households) and includes stakeholder-specific results, such as reduced conflicts among fishers and increased confidence in decision-making bodies. This specificity provides a more granular understanding of the impacts, which the AI content lacks in certain sections.

"

"57. Study Title : Influencing Organisational Strategy to Support Responsible Business Practice
Rater 1 Rating = 4 - Very Good
CR Justification: The AI-generated research impacts cover most dimensions and provide a well-rounded view across academic, social, environmental, and policy domains, including scalability, stakeholder engagement, and governance impacts. However, some of the community-specific and event-driven details are missing.

Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated content addressed the significant impacts across multiple domains, including economic, social, and environmental dimensions. It comprehensively covered the research's academic impacts, societal contributions, and cross-sectoral partnerships, aligning well with the REF submission. However, some of the depth around specific programs, such as Scotland CAN B and the Charity Board Initiative, was more detailed in the REF version.
Rater 3 Rating = 4 (Very Good)
CR Justification: The ChatGPT version covers key dimensions like social, economic, and technological impacts as described in the REF submission. However, a few intricate details regarding specific local organizations and their direct feedback were missing.
Rater 4 Rating = 3: Good
CR Justification: The AI-generated version covers significant impact areas discussed in the REF submission, especially social and economic impacts. However, a few impact dimensions such as detailed partnership dynamics and student involvement specifics were less emphasized.

"
"58. Study Title : Billmonitor: predicting the best mobile phone contract for businesses and individual users
Rater 1 Rating = 3: Good
CR Justification: The REF submission provides a comprehensive view of Billmonitor's impact, especially in terms of economic savings for various stakeholders (SMEs, NHS, and private individuals). It also includes detailed pathways to impact, such as media coverage, testimonials, and real-world applications. The AI-generated content captures most of the key points but lacks some specific examples and the depth provided in the real-world engagement details, such as endorsements by high-profile individuals and specific company savings figures.
Rater 2 Rating = 4 - Very Good
CR Justification: The REF submission provides an in-depth view of the economic impacts, partnerships, and direct user feedback, particularly on savings identified for businesses and consumers. It also discusses real-world implementations like partnerships with Ofcom. The AI-generated content covers the core elements but delves deeper into the theoretical framework and future potential applications across different disciplines. However, it lacks the same level of quantitative detail, such as the exact savings amounts and public endorsements present in the REF submission.
Rater 3 Rating = 4 - Very Good
CR Justification: ChatGPT's version covers multiple impact areas such as economic benefits, social savings, and healthcare sector influence. However, while the REF version offers richer detail on specific endorsements, user feedback, and quantitative savings estimates, ChatGPT emphasizes methodological and societal impacts more broadly. Both versions convey the foundational impacts of cost savings and algorithmic development effectively.
Rater 4 Rating = 4 (Very Good)
CR Justification: The AI report covers most impact dimensions discussed in the REF case, such as financial savings, improved decision-making, and the application of statistical modeling. Nonetheless, the original REF document includes additional coverage of the operational milestones and key stakeholders, providing a more exhaustive historical context on project development.

"
"59. Study Title : Mediating Modern German: reaching new and diverse audiences through translation, engagement, and performance
Rater 1 Rating = 3 - Good
CR Justification: The ChatGPT version covers most dimensions of impact, such as public awareness, translation work, and interdisciplinary collaborations, but lacks some of the more specific and detailed examples from the REF submission, especially regarding partnerships and personal engagements. While it captures the essence of public impact, it doesn't detail the depth of influence on competitions or media engagements.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover a broad spectrum, addressing several domains such as cultural impact, social engagement, and audience diversity. However, the comprehensiveness is slightly less detailed than the REF submission, particularly in terms of specific individuals, events, and accolades mentioned in the REF version. The AI impacts lack the depth of coverage regarding specific workshops, media appearances, and long-term collaborations with festivals and cultural institutions.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI-generated content addresses the primary dimensions of impact found in the REF submission, including social, cultural, and artistic impact, as well as aspects of educational outreach and engagement through workshops and public events. However, some specific

details—like the breadth of Leeder's collaborations and awards—were summarized more generally, leading to a minor reduction in comprehensiveness.
Rater 4 Rating = 4
CR Justification: The AI-generated impact covers most dimensions noted in the REF submission, including public engagement, policy influence, and broader cultural impacts. However, it lacks some details on event-specific engagement and collaborative networks Leeder developed with various institutions.

"
"60. Study Title : OMass Therapeutics: New technology for drug discovery with economic benefit to the UK
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover most significant dimensions mentioned in the REF submission, including technological, healthcare, and economic impacts. However, the AI version lacks depth in certain areas, particularly around specific company milestones, details of collaborations, and how the research led to the creation of OMass Therapeutics.
Rater 2 Rating = 4: Very Good
CR Justification: The ChatGPT version comprehensively covered the technological, economic, and academic impacts of the project. However, it was slightly less focused on specific industry collaborations and long-term investments that were well-documented in the REF submission.
Rater 3 Rating = 4 - Very Good
CR Justification: The AI-generated report covers most critical impact aspects from the REF submission, including technological advancements, the economic contributions through job creation, and the unique scientific methodology developed. The report aligns on economic, technological, and academic impacts but could deepen coverage on direct economic data like turnover and investment details.
Rater 4 Rating = 4 - Very Good
CR Justification: The AI content covered most impact dimensions, including economic, technological, and societal aspects. It also mirrored the strategic elements of OMass' growth trajectory. The original REF submission, however, provided more granular information on specific impacts like workforce development and detailed economic contributions, thus enhancing the comprehensiveness of REF's account.

"
"61. Study Title : Living With Feeling: Transforming Understandings of Emotional Health
Rater 1 Rating = 4
CR Justification: The AI-generated version covers most significant aspects of the research impacts, including academic, social, and policy domains. However, the ChatGPT version does not provide as rich a description of media engagement and public interactions as seen in the REF 2021 submission. The REF submission includes concrete metrics and feedback from stakeholders which add depth to the impact narrative.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated report covers key impact dimensions, such as public, academic, and societal impacts. However, it lacks the same level of detail and specificity found in the REF version, particularly regarding the metrics and quantifiable success of the public engagement initiatives, like the number of listeners, visitors, and specific stakeholder collaborations.
Rater 3 Rating = 4
CR Justification: The AI-generated impacts cover the project's broad social, cultural, and educational domains, capturing public education, healthcare, and policy domains. However, it omits some specific examples of program development and partnerships with nursing institutions that broaden the impact scope in the REF version.
Rater 4 Rating = 5 (Excellent)
CR Justification: ChatGPT's impacts cover all significant areas highlighted in the REF submission: social, cultural, educational, and public health influences. Its comprehensive scope includes the development of emotional health programs in schools and cultural engagement through public lectures, exhibits, and media. It also encompasses both immediate and long-term impacts, showing thorough understanding and application of historical and emotional concepts across sectors.

"
"62. Study Title : Litigation as a Tool to Support Social Change: Indigenous Peoples, Human Rights, and Legal Empowerment
Rater 1 Rating = 3: Good
CR Justification: The AI-generated impacts cover a significant number of relevant dimensions, including legal, social, and policy impacts. However, the coverage of specific legal cases and

certain stakeholders (e.g., local community leaders in the Namibian and DRC cases) is less detailed than in the REF submission. Some sections of the REF submission that deal with institutional impacts on NGOs like MRG and UN reports are more comprehensively covered.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated content comprehensively covers the major impact areas such as legal empowerment, social justice, and policy influence. However, some dimensions like specific contributions to UN declarations and local legal strategies in Africa are better elaborated in the REF submission.
Rater 3 Rating = 3: Good
CR Justification: While the ChatGPT version captures the primary impact dimensions (social, legal, and policy), it does not address all nuanced details included in the REF submission. Coverage on institutional impacts, like specific organizational responses to the research findings, and detailed impact on localized litigation outcomes, is less comprehensive.
Rater 4 Rating = 3: Good
CR Justification: The ChatGPT report covered multiple impact domains, aligning with most key themes in the REF report but lacking depth in some targeted areas, such as case-specific examples of litigation impacts across regions. This broader perspective is comprehensive but less exhaustive on specific community impacts than the REF case study.

"
"63. Study Title : Accelerating the development of medicines for children through an open-access excipient database
Rater 1 Rating = 4 - Very Good
CR Justification: ChatGPT provides a comprehensive coverage of the research's potential impact across several domains such as regulatory, societal, technological, and healthcare improvements. It addresses the main points found in the REF-submitted version. However, ChatGPT is slightly more general in covering areas like economic savings and specific excipient inclusions.
Rater 2 Rating = 3 (Good)
CR Justification: The ChatGPT version covers several major impact dimensions, such as economic, policy, and healthcare outcomes, but lacks some detailed, nuanced examples provided in the REF submission. The REF submission offers more specificity in terms of real-world applications and regulatory achievements, which the AI-generated version only covers in broader terms.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated version addresses each key dimension from societal to economic and regulatory impacts comprehensively. ChatGPT extends into potential future impact predictions, covering the multi-stakeholder involvement and long-term scalability of the STEP initiative, aligning well with the REF version's focus on the wide-ranging benefits for pharmaceutical companies and healthcare professionals.
Rater 4 Rating = 4 (Very Good)
CR Justification: The ChatGPT version covered most impact dimensions—policy influence, clinical practice, and economic impact—outlined in the REF version, demonstrating a thorough understanding of the STEP database's role. It did not, however, fully encompass the nuances in some stakeholder interactions, such as individual sponsor contributions or detailed savings data in drug development costs. Overall, the content maintained comprehensive coverage of major impact areas.

"
"64. Study Title : Pluralistic Evidence for Successful Policymaking about Reactive Systems
Rater 1 Rating = 3 (Good)
CR Justification: The ChatGPT version covers most key impact dimensions, including academic, societal, and policy impacts. However, it lacks some of the detailed narratives present in the REF submission, such as the ethical considerations and the depth of institutional influence on international organizations like IARC and NICE.
Rater 2 Rating = 4 - Very Good
CR Justification: The AI-generated version covers most of the dimensions of impact—academic, policy, societal, and economic—discussed in the original REF submission. It touches on the transformation in evidence assessment methodologies in health policy, technology, and ethics frameworks in data science. Nevertheless, it slightly lacks the depth in stakeholder involvement and post-publication engagement pathways highlighted in the original submission, particularly in shaping UK AI and data ethics frameworks.
Rater 3 Rating = 4 (Very Good)
CR Justification: The ChatGPT report captures the majority of significant impact areas covered in the REF submission, including healthcare, ethical AI, and public policy. However, while comprehensive, some nuanced elements of policymaker and practitioner collaboration are less emphasized in the ChatGPT version.
Rater 4 Rating = 4: Very Good

CR Justification: The AI-generated version covers multiple research impact dimensions—including social, ethical, and technological aspects—consistent with the study's aims. However, it lacks specific stakeholder interactions (such as direct consultations with NICE, IARC, and DEG) that were crucial to the policy adjustments mentioned in the REF case study. The AI version's broader coverage is notable, though it does not capture all the depth of each consultation detail.

"
"65. Study Title : Reshaping professional heritage practice and changing understanding of heritage in the UK and internationally
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI version addresses multiple dimensions of impact, including cultural, environmental, and policy domains, as seen in the REF submission. It also explores the long-term impact of heritage studies on sustainability and international policy. However, it misses the more detailed, concrete examples of specific initiatives, especially those related to professional development opportunities and co-design with stakeholders.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated version broadly covers most key aspects of the impact, including social, cultural, policy, and academic dimensions. However, it missed some of the granular details of specific interactions and events that the REF submission detailed. The AI version also tended to generalize some of the impacts that were more specifically described in the REF case study.
Rater 3 Rating = 4 - Very Good
CR Justification: The AI-generated impacts cover most of the dimensions discussed in the REF-submitted case study, such as social, environmental, and educational impacts, and the influence on public awareness. While comprehensive, the AI-generated report lacks some specificity, especially regarding examples of partner engagement activities and the quantitative reach (like audience engagement statistics) provided in the REF case study.
Rater 4 Rating = 3 (Good)
CR Justification: The ChatGPT-generated content covers the primary aspects of impact, addressing both policy and cultural domains and touching upon the role of exhibitions and community engagement. However, the ChatGPT content does not entirely capture the breadth of workshops and partner engagements specified in the REF case. The REF submission details various collaborative efforts, including professional development events, which deepen understanding in specific organizational contexts.

"
"66. Study Title : Shaping the legal framework for Brexit
Rater 1 Rating = 3 (Good)
CR Justification: While the ChatGPT version covers the major dimensions of the research impact—including academic, legal, and societal dimensions—it lacks the depth provided in the REF submission, particularly the intricate details about the parliamentary process and the specific legal citations used in the case studies.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated version captures many essential dimensions of the research impact, including policy, social, and legal domains. However, it misses some finer points related to specific citations and case law nuances that the REF submission includes. The REF submission delves more deeply into the historical and political context, such as discussions in Parliament and legal blogs that were instrumental in shaping public discourse.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated report captures a comprehensive range of impacts mentioned in the REF 2021 submission, covering judicial influences, policy impacts, societal implications, and academic contributions. The AI version also effectively addresses interconnected societal impacts, reflecting the comprehensive nature of the REF 2021 study.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated version adequately covered the primary impacts of the research, particularly in legal and political dimensions, such as judicial decision-making and public policy implications. However, the AI version lacked the same depth in discussing the research's broader societal implications, such as public engagement and international legal relevance, which were mentioned in the REF submission.

"
"67. Study Title : Bristol's materials research is keeping the UK's fleet of nuclear power stations safe and operating.
Rater 1 Rating = 3
CR Justification: The AI-generated content covers significant aspects of the research impacts, particularly in domains like academic, technological, and environmental impact. However, it

lacks some of the granular detail about the specific methodologies used, and the role of particular research programs (like ENVISINC and ENCAMP) in contributing to these impacts. The AI output also lacks depth in discussing the exact safety validation processes UoB supported for each reactor extension.
Rater 2 Rating = 4 (Very Good)
CR Justification: The ChatGPT version is comprehensive and covers all major impact dimensions. It delves into the academic, economic, environmental, and policy impacts, similar to the original REF submission. Both documents emphasize the importance of UoB's contributions to material science and safety regulations.
Rater 3 Rating = 5: Excellent
CR Justification: The AI-generated content addresses all significant dimensions of the impact, including economic, technological, sustainability, and safety contributions. It thoroughly discusses long-term applications and the broader implications of UoB's research on material degradation and energy security, reflecting a comprehensive understanding of both immediate and far-reaching impacts.
Rater 4 Rating = 3 (Good)
CR Justification: The ChatGPT report covers key impact dimensions like safety, sustainability, and long-term energy reliability. However, it lacks some technical specifics and does not include EDF-specific compliance needs detailed in the REF submission.

"
"68. Study Title : Radical advance in treating age-related macular degeneration leading to global impact in prevention of blindness
Rater 1 Rating = 3 (Good)
CR Justification: While the ChatGPT version covers the major health and economic impacts of the research, it lacks the depth found in the REF submission concerning specific national and international clinical guideline changes. The REF submission is more exhaustive, providing detailed examples of policy influence, including the NICE guidelines and global adoption of bevacizumab, which are only briefly mentioned in the ChatGPT version. However, ChatGPT does offer a more forward-looking perspective, discussing potential broader impacts, such as environmental sustainability and future healthcare savings.
Rater 2 Rating = 3 - Good
CR Justification: The AI-generated report covered significant dimensions of impact, including healthcare, economic, and policy impacts, reflecting the broad categories of the REF submission. However, it lacked some depth in specific examples from certain regions (e.g., India, Brazil) and international guidelines (like WHO recommendations). While these elements were generally covered, they were not as detailed as in the REF submission, which cited individual national decisions.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI-generated report encompasses major areas of research impact, including healthcare economics, clinical practice, and patient benefits, although it lacks some granularity in international legal aspects and policy references. Overall, it provides a thorough but occasionally less detailed summary.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover essential dimensions including economic, policy, and clinical implications. However, the REF submission more comprehensively details each step from local NHS impact to global policy changes, illustrating depth that is summarized rather than fully articulated in the AI version.

"
"69. Study Title : Rolling programme of research, centred on the National Joint Registry, to improve the outcomes of hip and knee replacements worldwide
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated report covers a wide range of research impacts such as healthcare policies, economic implications, and patient safety improvements. However, it lacks the depth provided in the REF submission, which goes into more detailed country-specific impacts, financial implications, and precise clinical follow-up recommendations. The AI-generated version addresses all dimensions but is less exhaustive in terms of specific examples and metrics, which are crucial for thorough coverage.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated content covers most major impacts from the REF submission, including health, economic, and policy domains. It also captures various dimensions of the research's broader influence, such as its scalability and potential for global application. However, some less immediate impacts, like perioperative mortality reductions and patient care improvements, were less thoroughly elaborated in the AI-generated text compared to the original submission.

Rater 3 Rating = 4
CR Justification: The AI-generated content covered most impact dimensions, including health, policy, sustainability, and implementation, which are similarly detailed in the REF submission. However, REF offers a richer account of the pathway of impact from initial study to global policy adoption, which the AI summary generalized more broadly.
Rater 4 Rating = 3
CR Justification: The AI-generated report covers major aspects of impact (health, policy, and economic), but lacks a few technical specifics mentioned in the REF case study. Notable gaps include direct mentions of long-term monitoring and targeted surveillance protocols emphasized in the REF.

"
"70. Study Title : Reducing breast and ovarian cancer occurrences in women at high risk
Rater 1 Rating = 5 (Excellent)
CR Justification: The AI-generated impacts cover a wide range of dimensions, including healthcare improvement, policy influence, public awareness, economic impact, and technological advancement. Both versions address the core areas of health, policy, clinical practice, and economic impact. The ChatGPT version added potential impacts on commercial genetic testing and cross-disciplinary collaboration, expanding the comprehensiveness.
Rater 2 Rating = 4 - Very Good
CR Justification: The AI-generated impacts address most major impact dimensions (social, health, and economic) that are covered in the REF submission. Both versions discuss the BOADICEA tool's role in improving healthcare decision-making, health outcomes, and clinical practice. However, the AI-generated version misses some specific social and policy impacts found in the REF version, such as patient organization collaborations and specific policy endorsements that solidify BOADICEA's influence on clinical practices worldwide.
Rater 3 Rating = 5
CR Justification: The AI-generated report provides a comprehensive overview of the BOADICEA tool's impact on multiple dimensions: health improvement, economic cost-savings, and educational outreach. All aspects discussed in the REF submission are mirrored, including preventive strategies, policy uptake, and international accessibility. The addition of future utility in evolving health frameworks enhances the report's comprehensiveness.
Rater 4 Rating = 4 (Very Good)
CR Justification: ChatGPT's version covers essential aspects across health, clinical practice, policy influence, and public awareness impacts as outlined in the REF submission, adequately summarizing most critical dimensions. It references the health and policy impacts but could further detail the specific global guidelines and institutional adoptions.

"
"71. Study Title : The Haydn Scale: Changing policy and practice for improving pupil behaviour in schools
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI-generated version is thorough in covering major impact areas, like teacher training, policy changes, and societal influence. Most significant dimensions are addressed, such as sustainability and cross-disciplinary relevance. A few specific impacts in the REF, such as feedback loops with educational institutions, are less emphasized in the AI version, but overall breadth is commendable.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI version covers most primary impact dimensions presented in the REF submission, addressing educational, policy, and social impacts. It appropriately highlights the social importance of improving classroom climate and the educational implications on teacher training, aligning well with REF's discussion on trainee feedback and training module adoption. However, it missed some intricate discussions, like specific testimonials and quantitative feedback from stakeholders that illustrate long-term engagement impact.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated version thoroughly covers all major dimensions, including policy, educational, and social impact, in addition to longer-term research benefits of the Haydn Scale. The AI effectively mirrors the diverse applications and impacts on stakeholders.
Rater 4 Rating = 4 (Very Good)
CR Justification: The AI-generated report broadly covers significant dimensions—policy, educational impact, and classroom practice. It comprehensively identifies stakeholder engagement and discusses the Haydn Scale's practical application in teacher education. However, some details, such as the extensive testimonial use and citation frequency from various educational stakeholders, which are detailed in the REF version, are summarized more generally in the AI report.

"

"72. Study Title : ""Definitive demonstration of the adverse cardiovascular effects of air pollution leads to policy change at the local, national, and international level""
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated content covers most impact dimensions but lacks some contextual depth seen in the REF, such as the role of specific policymakers and strategic pathways detailed for sustained impact (e.g., integration with WHO guidelines). The REF submission presents a step-by-step advocacy process that the AI does not fully replicate, although it does provide a general overview of policy impact and community engagement.
Rater 2 Rating = 3 - Good
CR Justification: The REF document is thorough in presenting a multidimensional impact on policy, media, healthcare, and public awareness, while the AI-generated version captures significant aspects of health impact but is less comprehensive in specific economic and media influence pathways. The AI version covers essential domains but is somewhat limited in addressing the full range of dimensions, such as named stakeholder engagements, which the REF outlines.
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated report covers most dimensions of the impact, including public health, policy, and environmental concerns. However, it lacks some depth in UoE-specific contributions and the detailed pathways to impact provided in the REF submission, such as engagement in UK policy development.
Rater 4 Rating = 3: Good
CR Justification: While the AI-generated impacts cover the main aspects of the impact, including health, environmental, and policy implications, certain specific impacts were not fully detailed, such as the incremental policy steps within the UK (e.g., Clean Air for Scotland strategy) and direct responses from entities like the British Heart Foundation (BHF) mentioned in the REF submission.

"

"73. Study Title : Navigating Inclusion in International Peace Processes
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers a broad spectrum of relevant impacts, including social, economic, and policy domains. It provides a comprehensive view of the database's applicability in real-world peacebuilding, conflict resolution, and gender inclusivity. However, the coverage lacks some detailed aspects noted in the REF version, such as the specific mentions of supporting women in Middle Eastern peace processes through programming and specific mentions of the PeaceFem app. The REF submission also places more emphasis on granular stakeholder engagement efforts and real-world training applications, which are not explicitly covered in the ChatGPT version.
Rater 2 Rating = 4 – Very Good
CR Justification: The AI-generated report covers most significant aspects from the REF submission, including social, policy, and technological impacts. The REF document, however, provides a more exhaustive account by including specific examples of PSRP's influence on particular peace negotiation stages and related training programs. Additionally, it discusses the PSRP's contributions to the OECD DAC Recommendation on the Humanitarian-Development-Peace Nexus in greater detail. While the ChatGPT version addresses these elements, it does so with less depth, resulting in a high but not entirely complete score for comprehensiveness.
Rater 3 Rating = 4 (Very Good)
CR Justification: ChatGPT's content covers most of the key domains described in the REF submission, including social, economic, policy, and legal dimensions. The generated impacts thoroughly include potential advancements in peace and conflict studies, gender inclusion, policy reforms, and cross-disciplinary impacts. However, there are gaps in specifics concerning the project's technological development (e.g., the PeaceFem app) and certain gender-related recommendations. These omissions prevent a rating of 5, despite a generally thorough coverage.
Rater 4 Rating = 5 (Excellent)
CR Justification: The ChatGPT version captures the multi-dimensional impacts outlined in the REF, including social, economic, and policy dimensions. The AI content extends the relevance of gender-focused peace processes to international and local scales, reflecting the REF case's depth in addressing both micro and macro-level impacts on policy, social welfare, and gender empowerment.

"

"74. Study Title : Documenting and Protecting Survivors of Torture and Ill-Treatment Living in Poor Communities
Rater 1 Rating = 5 (Excellent)
CR Justification: The REF submission provides an exhaustive account of the study's various

impacts across social, policy, and technological dimensions, covering each dimension explicitly. It discusses impacts on public awareness and NGO operations, supported by field-specific examples (e.g., workshops in Geneva and collaborations with NGOs in Tunisia). The ChatGPT version adds an extra layer by exploring potential future impacts (e.g., further use of digital tools and broader policy applicability) beyond immediate project results.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover the core areas of the REF submission, including social, technological, and policy dimensions of impact. The ChatGPT-generated content addresses social impact through community awareness and policy reform and includes the economic dimension by suggesting resource allocation toward torture documentation tools. However, it lacks detailed statistical outcomes present in the REF, such as app notifications and specific project milestones, affecting overall comprehensiveness.
Rater 3 Rating = 5 (Excellent)
CR Justification: ChatGPT provides a comprehensive analysis that covers various dimensions of impact, including social, policy, and advocacy changes. It reflects the multi-dimensional impact of the research in human rights documentation, with examples ranging from practical applications to academic influence, aligning well with REF's extensive case studies across countries.
Rater 4 Rating = 4
CR Justification: The AI-generated impact statement encompassed most major themes from the REF submission, including social, policy, and technological impacts. However, while it captured the broader context, some intricate details around specific impacts on public debate and advocacy activities were less extensively covered.

"
"75. Study Title : Transforming genomic selection in commercial breeding programmes for pigs, dairy goats, and poultry
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers significant impact domains such as economic, health, and environmental benefits in detail. However, the REF submission provides a deeper breakdown, especially in agricultural efficiency, sustainability, and animal welfare impacts, which is relevant for breeding programs. The AI impacts address the general dimensions effectively but would benefit from the same level of quantitative specificity that's in the REF version.
Rater 2 Rating = Rating: 3 (Good)
CR Justification: The AI-generated analysis effectively covers the main domains of the impact (economic, environmental, and social) described in the REF submission, though it lacks some of the financial and trait-specific details that contribute to a complete picture. The AI version provides broad coverage, discussing general breeding advancements and economic benefits, though with fewer quantitative details on stakeholder-specific improvements (e.g., annual increases in breeding stock productivity and associated profit gains). Overall, the AI's comprehensiveness in addressing the main domains is adequate, and it could be improved by including the precise metrics and year-over-year impacts that add depth to the REF submission's impacts.
Rater 3 Rating = 3: Good
CR Justification: The AI content covers most dimensions identified in the original REF submission, addressing economic, societal, and technological impacts. However, it lacks depth in financial and practical illustrations (e.g., precise percentage improvements for traits). The AI reflects the end-to-end benefit across industries but misses detailed financial quantifications crucial to the economic and societal aspects provided in the REF submission.
Rater 4 Rating = 3 (Good)
CR Justification: The AI version includes many significant areas of impact mentioned in the REF study, particularly economic and productivity improvements in livestock breeding, and addresses sustainability implications. However, the AI version does not fully capture the specific advancements for each targeted species nor the regional and policy impacts.

"
"76. Study Title : Employing polymer physics to improve gluten-free bread structure
Rater 1 Rating = 4 (Very Good)
CR Justification: The AI-generated version covers almost all primary impact dimensions, including economic, technological, social, and educational impacts, matching the depth found in the REF submission, particularly regarding commercial and industry applications. While both documents address a broad impact on polymer-based technologies, the REF version offers more specific figures on economic growth and product scope, such as the Genius brand's growth figures.
Rater 2 Rating = 3 (Good)
CR Justification: The AI content captures core aspects of impact, including technological innovation, economic influence, and consumer satisfaction. It covers the primary scientific underpinnings effectively but lacks coverage of secondary dimensions such as the educational

impact of mentorship within the university and the specific entrepreneurial ventures that were an outcome of the collaboration. The REF submission distinctly mentions multi-faceted outcomes that go beyond academic and industrial scopes, such as fostering a research-oriented entrepreneurial culture and supporting the broader University of Edinburgh's mission, which the AI report does not cover.
Rater 3 Rating = 5
CR Justification: The AI-generated version fully covers the wide array of impacts described in the REF submission, including aspects of product innovation, public and industrial impact, and the educational influence via mentorship. All major dimensions, such as economic and societal impacts, were addressed with substantial completeness.
Rater 4 Rating = 3 (Good)
CR Justification: While the AI version covers essential research outcomes and implications in the social and economic domains, it lacks depth in certain nuanced areas present in the REF submission, such as detailed product performance improvement and mentorship-driven entrepreneurship within UoE. The AI report could benefit from more explicit mention of the mentorship and masterclasses led by Genius Food's founder.

"
"77. Study Title : Peripheral Impressionisms: challenging perceptions of Impressionism
Rater 1 Rating = 3
CR Justification: While the AI-generated impacts provide a high-level summary covering economic, social, and cultural aspects, the REF-submitted impacts cover the depth and breadth of economic gains (e.g., increased art sales and museum sponsorship) more comprehensively. The REF version also delves into specific curatorial influences, shaping future exhibitions and educational events, aspects which the AI-generated impacts only lightly touched upon.
Rater 2 Rating = 3
CR Justification: While the AI-generated impacts cover most areas addressed in the REF submission, certain details are missing, particularly specific attendance statistics, detailed auction price metrics, and direct quotations from critics. This omission limits the full portrayal of the exhibition's widespread engagement and tangible economic impact, which were emphasized in the REF case study.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI version captures most impact dimensions, covering economic, social, and cultural impacts comprehensively, but lacks some specific exhibition statistics and in-depth financial details that underscore the market influence. It broadly outlines the exhibitions' impact, though missing specifics like the Musée des Impressionnismes' record attendance.
Rater 4 Rating = 3 (Good)
CR Justification: AI coverage includes major impact areas identified in REF—economic, social, and cultural. However, it does not cover the specific stakeholder interactions as extensively, nor does it specify financial figures, sponsorships, and quantitative attendance details.

"
"78. Study Title : Raising the profile of Scottish Literature through writing and consultancy
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated version addresses most of the cultural, educational, and social impacts present in the REF submission, such as Riach's public engagement through essays and exhibitions on Scottish literature. However, certain aspects, like public engagement metrics and Riach's deeper interconnectivity efforts between literature and visual arts, could be more comprehensively covered.
Rater 2 Rating = 3
CR Justification: The AI-generated report covers a broad spectrum of the research's cultural, social, and economic impacts. However, the AI version lacks some depth in the academic and linguistic-specific impacts emphasized in the REF submission, particularly regarding how each chosen aspect of Scottish literature informs cultural heritage and national identity. Moreover, while the REF version meticulously highlights the interconnection of art, literature, and national identity, this multidimensional approach is somewhat simplified in the AI report.
Rater 3 Rating = 3 (Good)
CR Justification: The ChatGPT-generated impacts address key cultural, tourism, and educational aspects; however, they omit some areas highlighted in the REF version, such as detailed descriptions of direct public interaction with The National newspaper articles and the Traveller's Guide distribution scale. While the AI version provides a thorough general overview, it lacks depth in capturing all dimensions such as the specific community reach of each initiative.
Rater 4 Rating = 4
CR Justification: The AI-generated impacts cover most significant areas outlined in the REF 2021 submission, including social, cultural, and educational dimensions. These were accurately

represented, capturing the project's multi-faceted approach to engaging the public with Scottish literature and supporting cultural tourism. However, certain specific details, such as exact distribution figures of publications and the depth of community engagement in exhibitions, were omitted or generalized in the AI version.

"

"79. Study Title : Secukinumab becomes the first interleukin-17A inhibitor approved for psoriatic arthritis
Rater 1 Rating = 5 (Excellent)
CR Justification: The ChatGPT-generated version addresses all critical dimensions of impact, covering academic, clinical, economic, social, and policy implications comprehensively. It goes beyond immediate outcomes, discussing the impacts on patient quality of life, healthcare costs, and long-term policy changes. This wide-ranging view provides a holistic perspective, in line with the REF submission but with additional foresight into areas such as ethical considerations and potential future advancements in treatment options.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated impacts broadly cover the significant research impact areas—social, clinical, and economic domains—but omit finer points of the original case study's details. For instance, the REF submission includes granular insights into the extended use of an autoinjector and the relative economic implications for Novartis within global markets, which are condensed in the AI version. This results in a generally comprehensive but slightly less thorough presentation of certain economic and technological dimensions.
Rater 3 Rating = 4 - Very Good
CR Justification: The AI-generated impacts cover most significant aspects of the REF impact, touching on clinical and patient benefits, healthcare implications, and economic impact. However, some aspects, such as regulatory approval specifics, are presented in more general terms, potentially affecting the depth of the comprehensiveness.
Rater 4 Rating = 4
CR Justification: The AI-generated version covers major impact areas, including clinical, economic, and patient-centered outcomes, similar to REF 2021. However, it does not encompass the specific study phases and multi-country regulatory journey as extensively. For example, the AI version discusses regulatory approval but lacks detailed descriptions of the EMA and FDA processes as in the REF submission.

"

"80. Study Title : Establishing the Facts, Developing Professionalisation and Enabling Transparent 'Pilgrim-centred' Communication in the UK Hajj Sector
Rater 1 Rating = 4 - Very Good
CR Justification: The AI version covered most impact dimensions, such as policy, economic, and cultural aspects, aligning well with the REF submission. The AI summary highlighted stakeholder collaboration (APPG, Council of British Hajjis) and targeted impacts in fraud prevention, regulatory frameworks, and policy influence across multiple levels, showcasing comprehensive coverage of significant domains.
Rater 2 Rating = 3 (Good)
CR Justification: While the AI version broadly covers all primary dimensions—policy, social, and economic impacts—it lacks the same level of detail across the various stages of the project highlighted in the REF version. Notably, the extensive co-production efforts, such as collaborations with the APPG and stakeholders for transparency, and the complex network of relationships within the Hajj industry, were more comprehensively detailed in the REF report.
Rater 3 Rating = 3 (Good)
CR Justification: The ChatGPT output broadly covers the primary impacts related to Hajj sector governance and professionalization, including fraud reduction and transparent communication for pilgrims. However, the depth of individual consultations and specific reactions from stakeholders like CBHUK and LHO are more comprehensively detailed in the REF submission, highlighting critical practical responses and stakeholder endorsements.
Rater 4 Rating = 5 - Excellent
CR Justification: The AI-generated version captures all major impact dimensions mentioned in the REF submission, including social, economic, and policy dimensions. It covers short- and long-term impacts, such as promoting transparency within the Hajj industry and facilitating dialogue among stakeholders. Additionally, the AI report includes reflections on the socio-cultural impact and ethical dimensions of transparency and consumer rights in religious travel sectors.

"

"81. Study Title : A Bridge for Spies: Overcoming the Practitioner-Academic Gap in Intelligence and Security

Rater 1 Rating = 4 - Very Good
CR Justification: The AI-generated impacts capture the significant dimensions of the study, including cross-disciplinary policy impact, societal influence, and economic implications. It covered domains such as policy, social, and educational impacts but lacked specifics about certain stakeholders like the Cabinet Office and Five Eyes as detailed in the REF. Additionally, the AI version provided a broader analysis of potential technological impacts.
Rater 2 Rating = 4: Very Good
CR Justification: The AI-generated content comprehensively covers the main impact dimensions highlighted in the REF submission, particularly policy, social, and academic impacts, and includes both short-term and potential long-term impacts. However, it does not fully capture all procedural or protocol-based changes, such as the NCA's open-source engagement changes and NATO's advisory role impacts, which the REF provides in more granular detail.
Rater 3 Rating = 3
CR Justification: The ChatGPT version covers multiple dimensions of impact, including social, economic, and policy implications, which are essential aspects of the REF-submitted case study. However, some specific impacts, such as the detailed training program outcomes and the influence on NATO's analyst program, are less thoroughly explored.
Rater 4 Rating = 3: Good
CR Justification: ChatGPT's analysis covers major impact dimensions such as security policy, professionalization, and collaborative frameworks with accuracy but lacks some depth in detailing specific partnerships. For instance, the REF version outlines specific milestones achieved through these partnerships, including analyst training across the Five Eyes and NATO. The AI-generated version touches broadly on these impacts but does not fully capture the extent of each specific agency's ongoing role in intelligence enhancement.

"
"82. Study Title : Influencing Labour Standards and Stakeholder Action Through International, European and National Law and Policy
Rater 1 Rating = 5 (Excellent)
CR Justification: The AI-generated impact report covers all essential dimensions outlined in the REF submission, including stakeholder impacts, domains of influence, and ethical considerations. It provides comprehensive insights across each research dimension, especially in areas like duration, stakeholder engagement, and scalability. The content's structure follows the intended scope comprehensively, representing both immediate and long-term implications in multiple geographical regions and policy contexts.
Rater 2 Rating = 4: Very Good
CR Justification: The ChatGPT version covers a wide range of research dimensions, such as social, economic, and legal impacts, with particular focus on austerity's effects on labor laws and collective bargaining. It also introduces cross-border transferability as a potential impact, indicating a scalable influence on other countries' labor policies. However, it lacks specific country-level details (e.g., Greece and South Africa) found in the REF submission.
Rater 3 Rating = 5
CR Justification: The AI-generated impacts comprehensively address the multiple dimensions covered in the REF submission, including economic, social, legal, and policy domains. AI content provides an extensive overview of the effects on social welfare, economic policy, and labor rights in crisis settings, capturing the broader and deeper ramifications as observed in the REF submission.
Rater 4 Rating = 3 (Good)
CR Justification: While the AI-generated content covers major domains (social, economic, legal), it lacks some critical details in scope, such as specific policy events and stakeholders. The REF submission provides richer contextual information on stakeholder actions, particularly detailing case-specific milestones and longitudinal outcomes across EU and national levels, offering a nuanced understanding of policy progression.

"
"83. Study Title : Creating Value and Transforming Lives through Arts and Creative Media Practice
Rater 1 Rating = 5
CR Justification: The AI-generated version provides a comprehensive view of the potential impacts across multiple dimensions, covering all significant aspects highlighted in the REF submission, such as social, economic, and educational impacts. It further explores sustainability, ethical considerations, and cross-disciplinary relevance comprehensively, mirroring the REF content. Both versions emphasize the long-term implications for urban planning, community capacity building, and arts-based interventions as a sustainable model for other urban areas.
Rater 2 Rating = 5: Excellent
CR Justification: The AI-generated impacts comprehensively cover social, cultural,

environmental, and policy dimensions of impact as per the original case study. Each significant area of impact, from heritage protection to community resilience and inclusivity in urban planning, is addressed, showing a nuanced understanding of the scope.
Rater 3 Rating = 3 (Good)
CR Justification: The ChatGPT version addresses most of the main aspects outlined in the REF submission, including social, policy, and cultural impacts. However, it lacks some granular details of event-based changes in policy and specific partnerships that led to significant outcomes. The REF version provides comprehensive insights into each impact pathway and the networks built within specific local communities.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated impacts provide a broad and accurate overview but do not fully cover all layers of the original studies. While social, economic, and cultural impacts are addressed, certain policy and stakeholder specifics, such as heritage preservation at particular sites, are underrepresented.

"
"84. Study Title : A Transformation in Creep Condition Monitoring for High Temperature, High Pressure Components
Rater 1 Rating = 4 - Very Good
CR Justification: The AI-generated version captures almost all significant impact dimensions, addressing areas such as technological advancement, economic value, and safety. However, the REF version provides added context in some instances, like the specific cultural and policy influence in South Africa, that contextualize the research more fully in specific industrial and regional impacts.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated impacts cover a broad range of impact dimensions, including economic, technological, and societal aspects. Most key areas, such as commercial and sustainability impacts, are well covered. The AI version, however, misses some specific mentions of societal impacts, like the South African recognition of innovation and the technology's social influence, which were emphasized in the REF version.
Rater 3 Rating = 3 - Good
CR Justification: The AI-generated impact analysis covers multiple dimensions, such as technological, economic, and social impacts, providing a well-rounded view. However, it does not cover the depth of stakeholder recognition or the complete detail of commercial and regional impacts as thoroughly as the REF submission. For instance, while the AI version discusses potential scalability in a global context, it does not delve into specifics about South African innovation awards or the exact economic benefits to eNtsa and SASOL.
Rater 4 Rating = 3
CR Justification: The AI-generated report covers critical aspects, such as the impact on industrial efficiency, technological advancement, and potential cost savings. However, it lacks comprehensive coverage of certain social and cultural impacts mentioned in the REF, particularly the acknowledgment of the recognition and awards the technology has received in South Africa, which underscores its social influence.

"
"85. Study Title : The creation (and re-creation) of contemporary female heroines at the center of new plays for the theatre.
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated version generally captures core impact areas, such as social change, gender representation, and cultural influence. It covers the overall dimensions such as public influence on gender roles in theatre and increased global representation for female playwrights. However, it lacks depth in certain specifics, such as the influence on UK-specific theatres and awards, and nuanced discussions of stakeholder-specific outcomes (e.g., direct testimonies from actors like Kirsty Stuart on the impact of roles for women in their careers).
Rater 2 Rating = 3
CR Justification: The AI-generated impacts cover core aspects, including public engagement, representation of women, and the play's wide reach. However, the REF version presented a more comprehensive breakdown across several dimensions such as programming practices at specific theaters, public debate impacts, and awards, demonstrating deeper coverage of the project's longitudinal cultural impact and international engagement.
Rater 3 Rating = 3
CR Justification: The AI-generated impact report covers the key dimensions of impact such as social, cultural, and policy influence, mirroring the original study's focus on gender balance and representation in theatre. However, it omits certain programmatic details that convey the full breadth of Harris's work's influence on international theatre and audience-specific feedback, reducing comprehensiveness.

Rater 4 Rating = [3 - Good]
CR Justification: The AI covers major impact dimensions (societal, cultural, and gender representation) and draws on multiple examples, matching the REF submission's thematic emphasis. However, it is not as extensive in detailing the individual impact of each play, audience responses, and international accolades.

"
"86. Study Title : Changes to cervical screening policies following the rollout of human papillomavirus (HPV) vaccination
Rater 1 Rating = 4 - Very Good
CR Justification: The ChatGPT version generally covers the health, policy, societal, and economic dimensions highlighted in the REF submission, with additional mentions of stakeholder and public awareness impacts. However, it lacks some of the detailed context and temporal milestones specified in the REF submission.
Rater 2 Rating = 3 (Good)
CR Justification: While the AI-generated impacts comprehensively address the significant public health, social, and policy dimensions mentioned in the REF submission, some economic and clinical impacts are less detailed. The original study emphasized the health economics of policy changes and subsequent financial implications for the NHS. The AI's coverage of technological, social, and policy domains remains robust, but omits depth in health economics and stakeholder-specific outcomes.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated content covers all major impacts mentioned in the original REF case study and also considers multiple dimensions—social, economic, and technological—demonstrating a comprehensive understanding of the study's implications.
Rater 4 Rating = 5 (Excellent)
CR Justification: The AI-generated impacts comprehensively cover all primary dimensions of the original study: policy change, clinical practice, and public health awareness. They effectively include both the social and health domains, capturing significant points on cervical screening policy and vaccine effectiveness. The ChatGPT version also addresses indirect stakeholders such as the general public, which adds depth to the comprehensiveness.

"
"87. Study Title : Combating Crop Losses and Improving Global Food Supplies through Mathematical Modelling of "Gene Silencing"
Rater 1 Rating = 5 - Excellent
CR Justification: The AI-generated content comprehensively covers all significant dimensions of impact from the REF submission, including environmental sustainability, economic gains, social impact on farmer practices, and policy implications. The AI report elaborated on implementation potential and scalability while providing detailed impact pathways, such as knowledge transfer methods and stakeholder engagement practices.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI report covers most dimensions discussed in the REF case, such as social, economic, and environmental impacts, alongside academic contributions. However, while the main aspects of yield increases and cost reductions are present, some details on the agricultural policy influence are missing. Both versions mention the economic and sustainability outcomes, but the REF offers more specificity regarding policy advocacy and regional social impacts.
Rater 3 Rating = 5
CR Justification: The AI-generated version comprehensively covers the significant aspects of impact across agriculture, environment, and economics, similar to the REF submission. Both versions addressed multiple dimensions, including short-term and long-term impacts, and presented impacts on public policy, food security, and farmer economic benefits.
Rater 4 Rating = 3
CR Justification: While the AI-generated content addresses major impacts—such as environmental benefits and agricultural sustainability—present in the original, it lacks certain dimensions. It does not capture the precise data-driven details of crop yield increases or cost reductions as well as regional statistics, which add depth to the REF-submitted impact statement.

"
"88. Study Title : Emerging Media, Learning, and Organisational Practice - Driving Change in Tourism and Education in Northern Ireland
Rater 1 Rating = 3 (Good)
CR Justification: The AI version captures most key areas—social, economic, educational, and technological dimensions—but lacks the depth of specifics found in the REF submission. For example, while the AI version mentions broad societal impacts and the economic benefits of AR

for tourism, it misses some of the nuanced, localized impacts on Northern Ireland's tourism initiatives and the detailed influence on specific educational institutions.
Rater 2 Rating = 3 - Good
CR Justification: The AI-generated version addresses a majority of dimensions such as economic, technological, and educational impacts highlighted in the original research. However, while it includes a broad overview of Jackson's contributions to app-based and AR-enhanced tourism, it does not fully encompass the multi-layered impact on regional policy-making and direct educational benefits reported in the REF submission. The REF version is more specific about the detailed reach and feedback from stakeholders, contributing to a slightly higher comprehensiveness in certain impact areas.
Rater 3 Rating = 4 (Very Good)
CR Justification: The ChatGPT impacts cover most dimensions highlighted in the REF, particularly in social, economic, and educational aspects. It includes comprehensive mentions of immersive technology adoption and its benefits across domains. However, nuances related to stakeholder-specific roles in Northern Ireland's tourism sector are slightly less detailed in the AI-generated version.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated impacts cover most significant aspects of impact in tourism and education, including innovations in mobile applications, immersive technologies, and pedagogical approaches. However, the comprehensiveness is somewhat limited as the AI-generated impacts do not fully capture the depth of Jackson's work in tourism strategies specific to the NI region or the exact educational impacts within individual schools. Additionally, the AI did not address broader implications on local tourism economy and specific stakeholder feedback that the REF submission includes.

"
"89. Study Title : Global adoption of the Dermatology Life Quality Index (DLQI) into clinical practice
Rater 1 Rating = 5 - Excellent
CR Justification: The ChatGPT-generated impacts provide a broad and comprehensive overview of DLQI's influence across healthcare, policy, patient outcomes, and economic impact. It captures both immediate and longer-term implications, touching on health policy, clinical decision-making, and technological adaptations. The AI-generated report's coverage across the diverse dimensions of impact showcases a thorough understanding of DLQI's contributions.
Rater 2 Rating = 3 (Good)
CR Justification: The ChatGPT-generated version provided an overview of DLQI's application across various domains, including healthcare, policy, and pharmaceutical use. However, it lacked some depth found in the REF-submitted impacts, such as the inclusion of specific regulatory bodies (e.g., NICE) and country-specific guidelines for DLQI use. The REF submission also included more detailed, quantitative information about the reach and specific usage figures for DLQI, which the AI-generated impacts lacked.
Rater 3 Rating = 4 - Very Good
CR Justification: The AI-generated report covers a wide range of impact dimensions, including social, economic, healthcare, and educational impacts, closely mirroring the detailed scope in the REF submission. However, while comprehensive, the AI's breadth of impact lacks certain case-specific details on cross-disciplinary reach that the REF submission emphasizes, such as application in other chronic diseases.
Rater 4 Rating = 3
CR Justification: The AI version addresses the main impact areas, including clinical adoption, economic revenue, and patient care enhancements. However, it lacks some depth regarding specific impact pathways detailed in the REF case, such as the extensive list of global guidelines and in-depth economic outcomes. The AI-generated content offers a comprehensive overview but does not match the granularity of the REF version in all dimensions, especially in economic and geographical impact mentions.

"
"90. Study Title : Improving the healthcare experiences of children and young people
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers a broad range of impacts from clinical practice improvements to educational and policy influences, aligning with many core themes from the REF-submitted version. Key areas such as health literacy, procedural anxiety management, and ethical considerations in pediatric care were well represented in both versions. However, the REF submission provided a more exhaustive account, especially regarding specific implementation pathways like targeted workshops, stakeholder engagement, and pre-and post-publication actions that promote longevity and breadth of impact in healthcare and education sectors. The AI version, while comprehensive, lacked the detailed granularity of targeted actions aimed at

ensuring sustained changes.
Rater 2 Rating = 3 (Good)
CR Justification: The AI-generated version adequately covers the significant dimensions of social, educational, and policy impacts, mirroring the areas explored in the REF submission. However, AI lacks some depth in detailing the pathways and collaborations that facilitated these impacts, reducing the level of comprehensiveness.
Rater 3 Rating = 3 (Good)
CR Justification: The AI-generated version covers the main impact dimensions, including health literacy and restraint policy influence, similar to the REF submission. However, the AI does not fully detail some key implementation pathways, especially the detailed feedback from stakeholders and specific contributions to national standards seen in the REF version.
Rater 4 Rating = 4
CR Justification: The AI-generated content covers the core areas of impact, including health, educational, and policy dimensions that were originally outlined in the REF case study. Both versions reflect a commitment to advancing pediatric healthcare practices and supporting families. However, the REF submission includes more detailed examples of specific stakeholder interactions, which were instrumental to its policy influence and are less specific in the AI version.

"
"91. Study Title : The value of the Carers' Alert Thermometer (CAT) in identifying family carers' needs and supporting them in their caring role
Rater 1 Rating = 4
CR Justification: The ChatGPT version comprehensively covers the main dimensions of impact identified in the REF submission, such as health improvements, policy influence, and technological adaptation. The REF version, however, includes several highly specific impacts, like CAT's use in training volunteers and specific adaptations in palliative care (e.g., MND Association and Stroke Association applications), which adds depth to its comprehensiveness.
Rater 2 Rating = 3 (Good)
CR Justification: The AI version adequately covers significant aspects of the REF impacts, including health and social implications, the CAT's adaptability to different carer demographics, and its relevance to clinical and community-based healthcare. However, while the REF submission integrates rich case-based evidence, especially testimonials and unique feedback from pilot implementations, the AI version lacks comparable granularity in stakeholder responses and feedback mechanisms.
Rater 3 Rating = 5 (Excellent)
CR Justification: The AI-generated impacts encompass nearly all significant aspects of the REF submission, covering health impacts, policy influence, and adoption in practice. It captures the tool's versatility in multiple settings and its contributions to healthcare policy. Additionally, the AI's discussion on the CAT's implications for clinical practice, mental health, and broader support in community healthcare settings reinforces the tool's wide application.
Rater 4 Rating = 3 - Good
CR Justification: The AI-generated impact report covers major dimensions of CAT's impact, including health, policy, and social aspects, but omits some specific geographical reach and detailed user feedback that enhance the narrative of CAT's broad adoption and feedback from real users.

"
"92. Study Title : God-complexity and the Multiple God-Aspects Framework
Rater 1 Rating = 4 - Very Good
CR Justification: The AI output covered nearly all key dimensions identified in the REF submission, including health, social, and educational impacts, suggesting a high degree of comprehensiveness. The AI-generated report expanded on cultural and societal impacts, though with less granularity on specific examples compared to the REF content.
Rater 2 Rating = 3
CR Justification: The AI-generated impacts broadly cover core dimensions such as societal, academic, and health impacts, reflecting the REF submission's interdisciplinary approach to God-complexity. However, while it touches on social cognition, psychological impact, and well-being, it misses the REF's detailed accounts of longitudinal studies and specific religious community responses, which demonstrate ongoing, specific impacts in applied and cultural contexts. Consequently, this version only partially encompasses the multifaceted impacts highlighted in the REF.
Rater 3 Rating = 4
CR Justification: The AI-generated content covers multiple aspects of the research impact described in REF, including social, psychological, and health impacts, as well as academic

contributions. However, the REF submission includes specific examples of conference presentations and pilot testing which add to comprehensiveness.
Rater 4 Rating = 3 (Good)
CR Justification: The AI-generated content covers major dimensions such as academic, cultural, and therapeutic implications, meeting the core impact categories outlined in the REF submission. However, it does not fully capture specific, documented impacts related to stakeholder interactions and the longitudinal aspect of the REF study. Thus, the AI report is well-rounded but could be expanded for detailed inclusivity.

"
"93. Study Title : Abertay Game Lab: play, performance, and public engagement with games
Rater 1 Rating = 3 (Good)
CR Justification: The AI report captures a substantial portion of the documented impact dimensions—particularly social, cultural, and industry-related impacts—by covering the primary stakeholders and achievements. However, it lacks some specific evidence and examples of direct testimonials, engagement metrics, and attendance data that were detailed in the REF submission. The comprehensiveness is adequate but could be improved by incorporating more granular data from stakeholder experiences.
Rater 2 Rating = 3
CR Justification: The AI-generated version addresses many significant impact dimensions outlined in the REF submission, including social, economic, and technological impacts on the gaming and arts industries. However, the REF-submitted version provides a more exhaustive account of beneficiary testimonials and specific institutional impacts on particular organizations. For instance, while both versions acknowledge the collaborative work with the V&A museum, the REF submission covers details about workshop attendance, exhibitions, and specific impacts on staff development within cultural organizations more thoroughly.
Rater 3 Rating = 3 (Good)
CR Justification: ChatGPT covered most major areas, including arts and cultural impacts, game curation, and industry development, but did not fully explore community-based impacts on specific stakeholders like local charities and practitioners. REF submission was more exhaustive, detailing specific stakeholder groups and partnerships.
Rater 4 Rating = 3
CR Justification: ChatGPT covers broad dimensions of impact across social, cultural, and economic aspects, aligning with the comprehensive nature of the REF submission. However, it lacks depth in addressing specific contributions to museum exhibitions and the academic outputs derived from each project. This affects its comprehensiveness, as it does not fully encapsulate all stakeholders or events.

"
"94. Study Title : Improving Care for Patients with Chronic and Distressing Tinnitus through Mindfulness Based Interventions
Rater 1 Rating = 5 - Excellent
CR Justification: The AI-generated content comprehensively covered the main impact areas, addressing clinical, social, and economic dimensions as well as detailed, domain-specific impacts such as public awareness and education. The AI version also addresses the broader application of MBCT for other chronic conditions and the potential for interdisciplinary research in psychology and audiology, which expands upon the REF version.
Rater 2 Rating = 3
CR Justification: The AI content covers most major impact dimensions, including clinical practice, healthcare policy, and patient well-being. However, while social and economic impacts are noted, these are less detailed, with minor gaps around specific economic metrics (e.g., cost savings data for healthcare providers) and the professional training impact mentioned in the REF submission.
Rater 3 Rating = 4 (Very Good)
CR Justification: The ChatGPT-generated impacts cover all major impact dimensions discussed in the REF submission, such as changes in clinical guidelines, patient satisfaction, training initiatives, and policy influence. However, some minor details and specific testimonial insights from healthcare providers are less emphasized in the ChatGPT version.
Rater 4 Rating = 4
CR Justification: The ChatGPT version covers core elements such as the healthcare impact, policy influence, and social acceptance of MBCT-t for tinnitus, reflecting the multi-dimensional impact. It highlights the therapeutic shift, economic benefits, and healthcare system integration. However, the report's depth could improve by including specific programmatic adoption rates and sustained patient improvement metrics, which are more elaborated in the REF document.

"95. Study Title : Developing and Embedding Effective Careers Guidance for Young People in England
Rater 1 Rating = 3 (Good)
CR Justification: The AI-generated content covers significant aspects of impact in terms of educational, social, and policy influences, capturing the core focus on career readiness and educational policy alignment. However, it omits some unique details regarding the career leaders' roles and training program statistics from the REF submission. It also lacks specifics on how international stakeholders and different educational systems are applying the benchmarks.
Rater 2 Rating = 4: Very Good
CR Justification: The AI-generated impacts cover all significant impact dimensions detailed in the REF study, including educational, social, policy, and technological impacts. The REF submission, however, goes into slightly more depth regarding certain contextual factors in career readiness, such as regional endorsement strategies. The ChatGPT version introduces interdisciplinary impacts, indicating how the guidance benchmarks might inform adjacent fields like mental health and social work, thus enhancing the perceived comprehensiveness.
Rater 3 Rating = 4 (Very Good)
CR Justification: The AI version covers the main dimensions, including social, educational, and policy impacts. While it touches on technology and cross-disciplinary relevance, it lacks depth in policy endorsements and specific examples of governmental support as seen in the REF submission, impacting its thoroughness.
Rater 4 Rating = 3
CR Justification: The AI-generated impact description covers major dimensions like policy influence, educational enhancement, and social benefits, but it lacks depth in certain impact areas such as specific policy-related guidance for career leaders and the geographic scope of influence. While the main points on social mobility and educational improvement are addressed, some critical REF-specific details, especially regarding Wales and Hong Kong adoption, were less pronounced in the AI version.

"
"96. Study Title : Design meets disability: changing the relationship between disability and design in business, culture, practice, and education
Rater 1 Rating = 5 - Excellent
CR Justification: ChatGPT's analysis thoroughly covers all significant aspects highlighted in the REF submission, including cultural, economic, technological, and policy dimensions. For example, both versions document the societal impact and shifts in consumer attitudes driven by Alleles, and both address educational aspects through Microsoft's inclusive design initiatives.
Rater 2 Rating = 3 (Good)
CR Justification: The AI version covers substantial dimensions, such as social, technological, and cultural impacts, as seen in the focus on inclusive design's societal relevance. However, it does not delve into all facets outlined in the REF submission, particularly around engagement activities and institutional collaborations. The REF submission provides more depth on the ways specific exhibits and collaborations impacted cultural understanding and institutional practices, adding breadth to the impact narrative.
Rater 3 Rating = 3 (Good)
CR Justification: The ChatGPT version covers significant dimensions, such as economic, technological, and cultural impacts, and aligns with the REF submission on most major themes. However, it lacks some of the specific examples, such as exhibitions and user testimonials that underscore cultural shifts in disability perceptions.
Rater 4 Rating = 4
CR Justification: ChatGPT's coverage includes key domains: societal, technological, and academic impacts, aligning with REF's multi-dimensional approach. However, while it captures the general impact, specifics, like the impact on Microsoft's Inclusive Initiative, were either missing or broadly generalized.

"
"97. Study Title : Digital Archiving for Curation and Dissemination
Rater 1 Rating = 5
CR Justification: The ChatGPT version covers all major impact dimensions identified in the REF submission, including cultural, academic, societal, and technological impacts. Additionally, ChatGPT effectively extends into implementation pathways like educational potential and public engagement, mirroring the extensive reach depicted in REF.
Rater 2 Rating = 3
CR Justification: AI covered most cultural aspects but lacked depth in social and economic impact dimensions presented in REF.
Rater 3 Rating = 4

CR Justification: The AI version comprehensively covers the primary dimensions of the REF submission, including the cultural, educational, and societal impacts, with an additional focus on the technological aspects of digitization. Both versions highlight the project's significance in making obscure archival content accessible, though the AI analysis introduces slightly broader applications in educational settings.
Rater 4 Rating = 3
CR Justification: The AI-generated content provides a good coverage of the essential dimensions of impact, including cultural, educational, and social aspects; however, it lacks the depth of REF's case studies, especially in demonstrating direct international partnerships and extended outcomes. REF's version goes beyond by contextualizing contributions to public policy and national archives and discussing their long-term societal impacts, particularly in cultural democratization.

"
"98. Study Title : Prevention and management of head injuries in cricket and rugby union
Rater 1 Rating = 3 - Good
CR Justification: The AI-generated impacts cover core dimensions such as helmet design, concussion management, and safety standards well, reflecting the REF-submitted impacts across most aspects. However, certain stakeholder impacts and programmatic updates were described with less precision. For example, specific community outreach efforts and injury-specific training programs for medical personnel were only briefly noted in the AI-generated impacts.
Rater 2 Rating = 4 (Very Good)
CR Justification: The AI-generated impact addresses most core aspects of the original submission, such as injury surveillance, helmet safety standards, and concussion management. It broadly covers the social, health, and policy impacts across cricket and rugby. However, some specific details were simplified, such as in-depth descriptions of partnerships and pathways to achieve safety standards.
Rater 3 Rating = 4
CR Justification: The AI report provides a comprehensive overview, covering domains such as health, policy, and technology, akin to the REF 2021 version. Both versions address impacts on stakeholders, though the AI lacks some specific figures and timelines found in the REF version. The AI additionally proposes applications in other sports and broader social implications.
Rater 4 Rating = 5
CR Justification: The AI-generated content covered nearly all essential aspects found in the original submission, providing a detailed perspective on both injury prevention and management across the two sports. This included the specific outcomes like new helmet safety standards and detailed policies within rugby concerning concussion management. Additionally, it encompassed aspects of both short- and long-term impacts on player safety, health, and regulatory policy.

"
"99. Study Title : The Kindertransport 1938/1939 to the UK: History Informing the Future
Rater 1 Rating = 5 - Excellent
CR Justification: The AI-generated version covers a comprehensive array of research impact dimensions, reflecting societal, educational, policy, and cultural impacts as detailed in the REF submission. Each key area from the original document, such as contributions to Holocaust education, public exhibitions, and impacts on policy debates around trauma-informed refugee care, is well represented. Additional dimensions, like ethical considerations and international relevance, are also included, providing a broader interpretation of potential future impacts.
Rater 2 Rating = 3 (Good)
CR Justification: The AI report addresses main impact dimensions such as educational, societal, and policy implications but occasionally lacks specific stakeholders and some depth of engagement details found in the REF report. It covers the general scope well but could benefit from a more detailed documentation of each public engagement instance and targeted stakeholder responses.
Rater 3 Rating = 5 - Excellent
CR Justification: ChatGPT's version broadly covers all significant areas mentioned in the REF submission, including public awareness, historical reflection on policies, and educational outreach. It effectively discusses the multidimensional impacts across social, policy, and educational domains and captures the full spectrum of influence from local public talks to broader policy implications.
Rater 4 Rating = 4
CR Justification: The AI-generated report covers many of the same dimensions found in the original case study, including public awareness, policy debate, and educational influence. Nevertheless, it misses out on some specific stakeholder feedback and a few niche historical contexts addressed in the REF version, especially regarding the UK's socio-cultural reception during Kindertransport events.

"

"100. Study Title : An evidence-based approach reduces the local costs of biodiversity conservation in low- and middle-income countries
Rater 1 Rating = 5 (Excellent)
CR Justification: The AI-generated impact report comprehensively addresses multiple dimensions of impact: social, economic, environmental, and policy. It covers local community influence, potential policy shifts, and broader socio-economic implications effectively, reflecting most facets seen in the REF submission. The AI-generated version also extends into potential for future integration across disciplines and highlights long-term influences on environmental justice in developing countries, adding depth to its coverage.
Rater 2 Rating = 5: Excellent
CR Justification: The AI-generated impacts covered all critical dimensions outlined in the original REF study, including environmental, social, and policy-related effects. The generated content effectively touched on various domains of impact (e.g., economic, societal, and ecological impacts) and discussed the implications of biodiversity offsets for local and international policy reform, thus meeting the criteria comprehensively.
Rater 3 Rating = 5 (Excellent)
CR Justification: The ChatGPT version provides comprehensive coverage across multiple impact dimensions (policy, community, corporate responsibility, and environmental sustainability). It includes short- and long-term impacts, the role of stakeholders, and the implications of biodiversity offsets for conservation and local communities. Additionally, it covers international relevance and the ethical aspects of offset practices, ensuring that most critical aspects are present.
Rater 4 Rating = 5 (Excellent)
CR Justification: ChatGPT comprehensively covers social, economic, and policy impacts mentioned in the REF submission, including cross-national impacts, which reinforces a full-spectrum impact understanding. The AI-generated version addressed social justice concerns by recognizing the economic impacts on vulnerable populations, aligning well with the detailed description of Bangor's influence on national conservation policies.

"